CHARACTERIZATION OF FUNCTIONAL DOMAINS OF A T-DNA PROMOTER ACTIVE IN SUNFLOWER TUMORS

By

WESLEY BERNARD BRUCE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1987

To the memory of my father, John W. Bruce

ACKNOWLEDGMENTS

I wish to thank Sabita Bandyopadyhay and Hampton McRae and their many assistants for their technical expertise. I also wish to thank Eva Czarnecka, Dulce Barros, and Ram Bandyopadhyay for guidance concerning many of the procedures involved in this study. I owe a debt of gratitude to John Ingersoll and Luis Mosquera for providing moral support as fellow graduate students when I needed it and for their unending humor, making lab work more enjoyable. A special mention goes to James Stanga for his brief and valuable help in screening some of the deletion mutants used in Chapter 2. I also owe a great debt of gratitude to my committee members, Robert Ferl, Frances Davis, James Preston and Curt Hannah, for their invaluable support and encouragement for this work and for providing solutions to my seemingly unending questions. I also wish to express my deepest thanks to William Gurley for support and guidance, for inspiring creative innovations, for his acceptance for new ideas and for simply being a friend in times of need. Above all, I would like to express my warmest thanks to my wife, Karen, for providing me with the courage and determination to continue with this work and seeing it to the end.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

ABSTRACT v

CHAPTERS

1. INTRODUCTION 1

Promoter Structure of Animal and Viral Genes 2
Plant Promoter Structure 15

2. PROMOTER MUTATION ANALYSIS 30

Introduction 30
Materials and Methods 31
Results 43
Conclusion 60

3. ENHANCER PROPERTIES OF T-DNA PROMOTERS 66

Introduction 66
Materials and Methods 68
Results 71
Conclusion 82

4. IN VITRO NUCLEAR FACTOR BINDING TO THE 780 ACTIVATOR ELEMENT 89

Introduction 89
Materials and Methods 90
Results 93
Conclusion 99

5. SUMMARY 106

REFERENCES 112

BIOGRAPHICAL SKETCH 121

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

CHARACTERIZATION OF FUNCTIONAL DOMAINS OF A T-DNA PROMOTER ACTIVE IN SUNFLOWER TUMORS

by

Wesley Bernard Bruce

May, 1987

Chairman: William B. Gurley
Cochairman: Francis C. Davis, Jr.
Major Department: Microbiology and Cell Science

Three promoter domains required for transcriptional expression of the 780 gene of T-right (pTi15955) were identified by mutagenesis. These domains are analogous to the TATA, upstream element, and enhancer typical of many animal and viral genes. Precise quantification of transcriptional activity in sunflower tumors of a series of 5' and internal deletions was achieved by using a double gene vector containing a reference 780 gene as an internal standard. Results of the 5' deletions delineated an activator element located -440 and -229 base pairs (bp) from the start of transcription. Removal of this element resulted in a 100-fold drop in activity relative to wild type. Large scale internal deletions (38 to 151 bp) occuring in various locations from positions -12 to -348 bp caused significant reductions in the major promoter transcriptional activity. However, internal deletions starting at position -37 and extending to –200 bp either had little effect, or increased activity. Removal of the TATA motif drastically reduced activity to <0.1% of wild type. The activator was shown to have enhancer-like properties by its position and its stimulation of

transcription in both polarities upstream of a -37 bp 5' deletion 780 gene even when positioned 538 bp further upstream than its normal location. However, the activator did not promote detectable levels of transcription when located 3' to the gene. This element was also shown to specifically interact with nuclear factors present in sunflower and soybean as determined by electrophoretic mobility assays and DNase I protection analysis. The upstream sequences (ca. –330 to -120 bp) of the octopine synthase (OCS) and agropine synthase genes of T-DNA did not promote efficient transcription when inserted in either polarity, upstream or downstream of the deleted 780 gene with one exception. The OCS gene fragment stimulated transcription to 15% of 780 wild type activity level while in the reverse polarity upstream of the deleted gene. In conclusion, the 780 gene appears to have a complex promoter structure similar to animal and viral genes in spite of its bacterial origin and constitutive expression.

CHAPTER 1

INTRODUCTION

Historically, procaryotic promoters provided the first evidence of *cis*–acting DNA sequences controlling gene expression. Bacterial promoters have been characterized as regions of DNA that bind RNA polymerase for specific initiation of transcription (39). The sequences of a multitude of procaryotic genes are known and generally show two highly conserved regions 5' to the start of transcription that are involved in control (143). One of these regions, labeled the "Pribnow box" or the "TATA box," has a sequence consensus of 5'–TATAATG–3' located at position –10 basepairs (bp) from the start of transcription. The other control region, known as the "recognition site," is approximately located 35 bp upstream of the start site. Other sequences upstream of the recognition site confer a negative or positive regulation in some genes. Highly regulated transcription usually entails either repressor proteins binding to operator DNA preventing RNA polymerase from binding, or an activator protein stimulating transcription by binding to specific DNA sequences farther upstream and presumably making contact with the RNA polymerase (62).

Based on sequence comparisons, eucaryotic promoters show some striking similarities to procaryotic promoters mainly of the "Pribnow box" and the "recognition site." The conservation of sequence homology in the 5' flanking region of genes suggests a preservation in the basic mechanisms of transcriptional control between most organisms. Functional analyses of these

highly conserved regions, as well as other sequences possibly required for eucaryotic gene expression, can be made by in vitro mutagenesis and using both in vivo and in vitro transcription systems. Previous testing of promoter mutations of animal and viral genes has established domains that govern transcriptional regulation. These domains have roughly been catagorized into three major elements based on sequence, location and function (13, 35, 138).

Promoter Structure of Animal and Viral Genes

The TATA Box

Of the promoter domains of animal and viral genes that are similar to functional domains in procaryotic promoters, an AT–rich region seems to be the most highly conserved. This region contains the sequence known as the TATA box and is usually found approximately 21 to 35 bp upstream from the start of transcription. Most of the genes sequenced show homology in this region with a consensus of 5'–TATA($^{A}/_{T}$)A($^{A}/_{T}$)–3' (13). The TATA box functions simultaneously in positioning the start of transcription and maintaining the rate of transcription (35, 154). Using deletion mutants of the conalbumin and the adenovirus–2 (Ad2) major late genes, this motif was shown to be required for accurate initiation of transcription (21, 160). Additionally, the TATA region alone from the Ad2 major late gene (–12 to –32 bp) initiates specific transcription starting about 25 bp downstream even when cloned in the plasmid pBR322 (131). Removal of TATA sequences from the simian virus 40 (SV40) early region (97) as well as from the sea urchin histone H2A gene (54) does not eliminate activity but produces heterogeneous sites of initiation. However, a single base pair transversion (T to G) at the second T in the conalbumin TATA box drastically reduces transcriptional activity in vitro

(159). A similar result occurs when point mutations are made in the TATA motif of the mouse β_{major} globin gene introduced into HeLa cells which reduces transcription to 25–50% relative to wild type levels (112). The same point mutations in the globin gene also altered the initiation site. In another study, Tokunaga et al. (154) demonstrated in vivo the requirement of the TATA box for specific transcription in the *Bombyx mori* fibroin gene expressed in monkey COS cells. An internal deletion (–58 to –20 bp) and point mutations in the TATA motif drastically affected the transcriptional activity while altering sites of initiation. Apparently, the TATA box only functions in positioning the start of transcription for some genes. In other cases, the TATA is required for normal transcriptional activity as well in the determination of the initiation site.

Although the eucaryotic RNA polymerases have not been shown to bind directly to specific regions within the promoter (90), other cellular factors are able to complex with DNA in the TATA motif and at other 5' control sequences. The binding of these cellular factors in turn may possibly provide the sites for eucaryotic RNA polymerase II to interact with the promoter in a specific fashion. The TATA boxes in the conalbumin promoter and in the Ad2 late promoter form stable pre–initiation complexes with a HeLa cell factor in the absence of RNA polymerase II (25). Parker and Topol (121) also demonstrated binding of a chromatographically distinguishable component prepared from isolated nuclei of cultured *Drosophila* K_c cells to the TATA–proximal regions of the *Drosophila* histone H3, H4, and 5C genes. Footprint analysis on this concentrated 70 kD "B factor" revealed the presence of sequence specific DNA–binding activity to regions which included the TATA box, the start of transcription and a portion of the leader sequence on these histone genes. The TATA sequence from a *Drosophila* heat shock gene, *hsp70*, can be protected

from exonuclease III digestion between –21 to –41 bp relative to the cap site with a factor present in *Drosophila* cell nuclei (171). The author of this latter study believes that this factor involved with the *Drosophila* heat shock gene may bind selectively to heat shock TATA boxes and not to analogous motifs of other genes which are transcriptionally inactive during heat stress (171). Based on this assumption, TATA–binding functions may exist as a family of proteins each recognizing unique sequences surrounding the TATA motif.

Based on sequence analysis, a few animal gene promoters seem to deviate from the requirement of having a TATA box. Instead of having this motif, these promoters have GC–rich sequences similar to the GC box of the SV40 upstream element (see below). These unique promoters direct transcription of genes generally involved with purine and pyrimidine metabolism such as hypoxanthine phosphoribosyl transferase (104), mouse dihydrofolate reductase (dhr) (101, 132), and hamster hydroxymethylglutaryl–coenzyme A reductase (124) and are generally thought to be constitutive in their expression. Since these genes lack the TATA motif, an alternate method of positioning the start of transcription may exist possibly involving the GC sequences. The GC–rich sequences of the dhr gene also interact with specific factors which may play a role in specifying the site of transcriptional initiation together with maintaining the constitutive level of expression (33, 40).

The Upstream Element

Based on promoter mutation studies and the presence of conserved groups of sequences, a region of the 5' flanking sequences in animal and viral genes between –40 to –110 bp from the start of transcripton is defined as the "upstream element" (30, 103, 112). Five different consensus sequences,

including the CAAT box, have been found to be individually present within this element (38). In addition, various promoters appear to have different combinations of one or more of these subelements present within the upstream element. Many genes require the upstream element for active transcription since deletions or point mutations located within this region often results in a decrease in transcriptional activity (36, 52, 59, 103, 112, 162). However, point mutations located immediately upstream of the CAAT box in mouse β–globin (112) and Herpes simplex virus thymidine kinase (HSV tk) promoter (52) result in a dramatic increase in activity, sometimes as much as 3.5–fold. These particular up–mutations may act by optimizing the affinity of a transcription factor for this element, or by possibly altering the conformation of this factor when it is bound to the promoter (52).

The upstream element is not always required for wild type activity of the promoter since some genes may be expressed at normal levels, or higher, if the upstream element is removed leaving TATA intact. For example, deletion of a region from –116 to –61 bp, that includes the CAAT sequences, from the sea urchin H2A histone gene injected into *Xenopus* oocytes increase activity nearly 2–fold (54). In another study where the 5' sequences of the fibroin gene were removed to position –44 bp with only the TATA motif remaining, normal rates of transcription in vivo (154) were still observed. Removal of TATA sequences, however, resulted in a complete loss of transcriptional activity in this gene. The upstream elements of the human β–globin gene can also be deleted resulting in normal transcription levels when transferred into cells that express the Ad 2 viral E1a gene products (53). The effects of transcriptional activation of the β–globin gene by the E1a gene products seem to be mediated through the TATA box, since the β–globin gene does not require the upstream element for activity.

In genes requiring the upstream element for activity, spacing between TATA and subcomponents of the upstream element is somewhat inflexible for optimal promoter expression. For example, the promoter of HSV tk gene contains three regions important for transcription (102). Two of these regions constitute the upstream element and are designated ds–1 and ds–2, while the third region designated ps includes the TATA box. When 6 bp are removed between ps and ds–1, the level of transcription drops to 10–20% of wild type activity. However, 9 to 30 bp can be inserted between ps and ds–1 without adversely affecting transcription. Increasing the distance between the upstream element and TATA also reduces promoter activity level, since the introduction of more than 30 bp results in less than 10% activity. The spacing between ds–1 and ds–2 is much less flexible in that only 10 bp can either be inserted or removed between these two sequence motifs without resulting in a decrease in activity. Stringent spatial requirements are also seen in the organization of the upstream element of the rabbit β–globin promoter in which only 10 bp insertion or deletions between two upstream element consensus sequences (the 5'–GCCACACCC–3' positioned at –90 bp and the CAAT box at approximately –75 bp) are tolerated without substantially affecting promoter function (30). It can be argued from these results that these three subelements of the HSV tk and the rabbit β–globin promoters bind factors and allow the factors to interact with each other directly, since removal or insertion of sequences between the binding sites would either crowd or separate these factors from optimal contact and eliminate activity.

Cochran and Weissmann (19) examined the interchangeability of upstream elements and TATA boxes between rabbit β–globin and HSV tk promoters. Fused tk–β–globin promoters were analysed by quantitative S1 nuclease mapping. Activities of mosaic promoters which included the β–globin

TATA and the tk upstream element had relatively high activities. Promoters that included the tk TATA sequence with the β–globin upstream element alone or together with the tk upstream element, however, resulted in low expression levels. Based on these results the β–globin TATA sequence seems to show no particular preference for upstream elements for maintenance of normal transcription levels. The tk TATA sequence may otherwise be of a specific nature and require its own upstream element to achieve wild type levels of activity.

Like the TATA box, the upstream elements of many genes also interact with specific nuclear factors (35, 49, 58, 71, 112). The HSV tk, human β–globin and mouse α_1–globin gene CAAT boxes interact specifically with a cellular factor present in nuclear extracts of HeLa cells, and at least for the tk promoter, the factor stimulates in vitro transcription (20, 34). A HeLa cell factor protects from DNase I digestion a specific region from 16 bp upstream to 9 bp downstream of the HSV tk CAAT box and 17 bp upstream to 6 bp downstream of the murine sarcoma virus LTR CAAT box (52). Another previously purified transcription factor known to interact with a subelement of the upstream element, the GC box, is the Sp1 factor obtained from cultured human cells (34). This factor binds specifically to the hexanucleotide sequence 5'–GGGCGG–3' (GC box) which can be found in the upstream elements of several genes including the SV40 early promoter (34), the HSV early gene 3 (70), the HSV tk gene (71), the human metallothionein–I_A gene (125), and the rat type II collagen gene (80). The Sp1 factor has been shown to be required for specific in vitro transcription of the SV40 early gene (34). The GC box of the chicken δ–crystallin also seems to bind specifically to an Sp1–like factor in vivo (58). Expression of the δ–crystallin promoter can be reduced when it is coinjected into mouse lens epithelial cells with either the sequences of the SV40 21 bp

repeats, or with the part of the promoter of the HSV tk gene that includes the upstream element. This reduction in promoter activity is presumed to be due to competition for nuclear factors (58). Clearly, the upstream element and the TATA box determine transcriptional levels by interaction with specific nuclear factors.

The Enhancer Element

The third subdivision of eucaryotic promoter elements, designated enhancers, (6, 160) are usually 100–200 bp in size and stimulate transcription of heterologous promoters. These promoter elements function bidirectionally and can operate over a considerable distance. This enhancer effect was discovered by two groups, Banerji et al. (4) and Moreau et al. (108) and then later seen by Fromm and Berg (46). As an example of enhancer activity, the far upstream sequences of SV40 stimulate transcription of a linked β–globin gene in vivo by more than two orders of magnitude in both polarities and over distances of more than 3000 bp, even from a position downstream of the gene (4). Transcriptional control elements with the same characteristics were later found in other viruses and cellular genomes (29, 93, 94).

The SV40 enhancer is the best characterized of the many known enhancers. This enhancer element contains two separate domains, "A" and "B," which are within a 72 bp repeat that is located in the promoter/regulatory region of the early and late genes (176). The 72 bp repeat sequence contains both a core consensus homology found in other viral and cellular enhancers consisting of the sequence 5'–GTGG($^{A}/_{T}$)($^{A}/_{T}$) ($^{A}/_{T}$)G–3' (163) in addition to stretches of alternating purines and pyrimidines (115). Maximal enhancer activity is acheived when domain "A" is stereospecifically aligned with the 21 bp repeats of the SV40 upstream element suggesting that the different factors

present along one side of the DNA helix interact by direct contact (151). This enhancer also acts in a constitutive rather than a tissue–specific or developmentally controlled manner in a wide variety of animal tissues and hosts (84, 85, 115).

The SV40 enhancer has been shown to activate a variety of heterologous promoters (4, 46, 108, 156) and will also initiate transcription without the apparent presence of any other promoter element (42). Insertion of the 72 bp repeat in either orientation immediately upstream of a conalbumin promoter (fused to the SV40 early gene coding region) promotes transcription starting at the cap site (161). Interestingly, when a 275 bp fragment from pBR322 is inserted between the SV40 enhancer and the –102 to +62 bp conalbumin promoter, transcription is intiated at a 27–31% activity level within the pBR322 sequence 40 bp downstream of the enhancer. In this construction, the SV40 enhancer also initiated faithful transcription at the cap site of the conalbumin promoter, but only at an 8% activity level. The pBR322 sequence upon inspection does not show any TATA–like motif, suggesting that the SV40 72 bp repeat can initiate RNA transcription from TATA–dependent as well as –independent start sites.

The SV40 enhancer demonstrates preference for the most proximal promoter when it is present on the same DNA molecule with more than one gene. Preferential activation of proximal promoters is demonstrated by the placement of two conalbumin promoters tandemly downstream of the SV40 enhancer (161). The enhancer activates transcription of the proximal conalbumin promoter to 60–80% activity while the distal promoter is only stimulated to 4–7% activity when compared to a single conalbumin promoter downstream of the enhancer. Kadesch and Berg (72) also observed similar results by inserting the 72 bp repeat of SV40 in various locations relative to

three distinct coding regions fused to truncated SV40 promoters. One of these coding regions, the xanthine–guanosine phosphoribosyl transferase gene (XGPRTase), was assayed for activity in CV–1 cells to determine positional effects of the SV40 enhancer on this gene. With the enhancer element in the 5' position, the proximal fusion gene was transcribed efficiently, whereas the activity of the distal gene, present further downstream, was reduced. Adhya and Gottesman (1) also observed this effect and named it "promoter occlusion." They suggested that the RNA polymerase II transcribing through the downstream promoter may interfere with the promoter's activation.

Another phenomenon designated as "enhancer dampening" is observed when the SV40 enhancer is positioned between two tandem genes which are in the same polarity. The level of activity of the upstream gene (XGPRTase) drops 3–fold compared to the activity observed when the XGPRTase gene is present with the enhancer sequence downstream alone (72). A possible explanation for these effects of promoter regions preventing the enhancer from affecting distal promoters is that these promoter regions might block a bidirectional movement of RNA polymerase II or other transcriptional factors binding to the enhancers (72). Another explanation is that the promoters of these two genes compete for the same set of enhancer related factors.

Some enhancer elements can function in a tissue–specific manner, such as those found in the mouse and human immunoglobulin heavy (38) and κ chain genes (7), the rat chymotrypsin (158), and insulin genes (158), the human type I keratin (96) and albumin genes (119), and the long terminal repeats of the Maloney murine sarcoma virus (85). A region of the insulin promoter from –100 to –400 bp relative to the cap site promotes the expression of the SV40 T–antigen only in the pancreatic islets of Langerhan β cells (158). The enhancer of the mouse polyoma virus, a papovavirus like SV40, shows

distinct host–cell preference being approximately four times as active in mouse cells than in primate cells, unlike the SV40 enhancer which promotes transcription equally well in both hosts (28). The immunoglobulin enhancers also direct transcriptional activity in lymphoid B cells early in differention, but once established, the enhancer sequence becomes dispensible (79). Enhancers as a class, therefore, seem to exert a continuum of specificity with regard to function, from the constitutive action of the SV40 enhancer to the highly specialized activation of transcription by enhancers associated with genes under stringent developmental and enviromental control.

Some promoter elements do not seem to fit neatly into the classes discussed so far. Promoter elements possessing some enhancer–like qualities of bidirectionallity and limited spacing flexiblity exist in several inducible genes such as the mouse metallothionein gene (MT–1) (74), heat shock genes (76), and the β–interferon gene (47). These enhancer–like elements are found in the region regarded as the upstream element as well as in various TATA–distal positions. Searle et al. (135) were able to demonstrate that promoter strength depended on the number of copies of the metal responsive element (MRE) sequences present in the 5' flanking region of a heterologous gene. They placed the MT–1 MRE sequence (5'–CCTTTGCGCCCG–3') in various locations and polarities within the HSV tk promoter and examined the regulation of induction by the addition of zinc. At least two MT–1 MRE's were required regardless of position or orientation to obtain low induction, but this induction was significantly increased when the elements were placed in the TATA–proximal position. The level of inducibility could be increased by incorporating more copies of this element into the tk promoter. In another study, heat shock consensus elements (HSE) of the *Drosophila hsp70* gene were inserted in multiple copies into the promoter of a heme–inducible yeast gene

(*CYC1*, iso-1-cytochrome c [88]) increasing the level of heat inducibility greater than 100-fold (162). The orientation of the HSE's in *CYC1* promoter did not affect the overall level of heat induction but adding more elements increased the inducible activity. Although MRE's and HSE's have the ability to activate heterologous promoters and function bidirectionally, they do not efficiently promote transcription while 3' to the gene.

Enhancers often seem to utilize a repetition of sequence motifs to activate transcription. Although enhancer elements of various viral and cellular types do not show any obvious repeats (29, 85, 139), obscure sequence redundancy may be present. A 74 bp region of the SV40 enhancer which includes portion of one of the 72 bp repeated sequences as well as some nonrepetitive upstream sequences can activate the SV40 early gene transcription (42). However, within this 74 bp region there are several shorter sequences which may provide the enhancer with necessary repetitive domains to promote high gene expression levels. Deletions of one of these shorter repeats does indeed eliminate activity. The SV40 enhancer can also activate transcription even more efficiently when the 74 bp sequence is repeated and in the presence of the upstream nonrepeated sequences (42, 176). To demonstrate the affect of duplicating sequences on activity, a truncated SV40 enhancer having reduced stimulatory effect was dimerized in vitro and transfected into COS cells resulting in restoration of transcriptional activity (28, 85). In another example, severely deleted enhancers fused to the chloramphenical acetyltransferase (CAT) coding region were introduced into COS cells and under selective pressure with chloramphenicol the deleted enhancers were duplicated in vivo (61). This in vivo duplication of the enhancer region in response to selective pressure suggests that the redundancy of this element plays a role in normal enhancer function.

Just as the other eucaryotic promoter elements were shown to bind specifically to transcriptional factors, enhancer sequences were likewise demonstrated to interact with *trans*–acting factors. Since some enhancers mediate tissue–specific and sometimes host–specific control, cell–specific factors seem to be involved. Schöler and Gruss (134) demonstrated factor interaction with enhancer DNA of an SV40/CAT fusion gene using a competition assay. They reported a reduction in activity of this gene when competed with normal SV40 enhancer sequences, but found no reduction in activity when competed with DNA fragments containing the 21 bp repeats, the TATA box or transcriptional termination signals of SV40. This indicates that the enhancer specifically interacts with cellular factors which are required for the activation of the SV40/CAT gene and that the presence of this factor(s) is in limited amounts within the cell.

Other groups have reported transcription factors binding to cellular enhancers such as those from the immunoglobulin heavy genes (IgH) and the κ chain genes (38, 137, 147). Mercola et al. (105) demonstrated by using competition assays that factors from lymphoid B cells bound specifically to the IgH enhancers, and that in vivo these enhancer sequences could compete with the SV40 enhancer for *trans*–acting factors. Common or closely related factors seem to interact with both of the IgH and SV40 enhancers since these enhancers have homologous sequences. Extracts from lymphoid B cells protect different regions of the mouse IgH enhancer DNA from in vitro DNase–I digestion when compared to extracts from HeLa cells (3). The mouse IgH enhancer can also efficiently activate the Ad2 major late promoter in vitro in the presence of lymphoid B cell extracts. These results suggest that the lymphoid cellular factor(s) binds to different regions of the IgH enhancer than the HeLa cell factors.

Heat shock elements also interact with a specific transcription factor which is activated when thermally induced (121, 155, 170). Based on exonuclease protection analyses, a region from –91 to –52 bp of the *Drosophila hsp82* gene is protected from digestion when in the presence of extract from heat–shocked *Drosophila* cells (169). The *Drosophila* heat shock transcription factor (HSTF) binds to three domains upstream from the TATA on the *hsp70* gene (170). Two of these binding sites occur within the region from –100 bp to the cap site and coincide with HSE's. Both HSE's are required for in vitro and in vivo transcriptional activation (162). This factor first occupies the TATA–proximal HSE which in turns facilitates the cooperative binding of a second HSTF to the TATA–distal HSE. The cooperative binding to the second site has been suggested to serve as the molecular switch that activates the *hsp70* gene (32). Shuey and Parker (141) demonstrated that a subset of protein–DNA contacts between the HSTF and the first HSE changed upon the binding of a second HSTF to the neighboring HSE. This change in the protein–DNA contacts suggests that a conformational change in the protein–DNA complex occurs. Part of this conformational change may involve bending of the DNA by interaction with the second HSTF protein (142). This bending of the DNA, upon binding of a regulatory protein, may be a general phenomena in the interaction of dimer proteins to DNA, since the catabolite activating protein of *Escherichia coli* also causes a bend or kink in the *lac* promoter DNA (172).

To summarize, three major subdivisions of promoter elements in animal and viral genes have been discussed and their specific function examined. The TATA element has an important role in governing the start of transcription and is essential in most cases for normal activity. The second region of a typical promoter, the upstream element, is also essential for activity in most genes. This element must act in close proximity to the TATA box, and with some

examples the subelements within the upstream element can be oriented in either direction. The enhancer comprises the third class of regulatory components within eucaryotic promoters playing an active role in establishing tissue–specificity, cell–cycle control or simply constitutive expression. Enhancers have the ability to activate heterologous promoters in either polarity, 5' or 3' to a gene, over large distances. A class of enhancer–like elements exist in many inducible genes having a combination of characteristics from the upstream element and enhancers. These elements, unlike true enhancers, require specific TATA–proximal locations (at least for one of the repetitive elements in a promoter) but can activate heterologous promoters in either orientation. The sequence elements responsible for heat shock and heavy metal activation fall into this intermediate class, and seem to activate transcription best in multiple copies. Specific nuclear factors also play an important role by interacting with all of these promoter elements and controlling the transcriptional machinery in a very specific manner.

Plant Promoter Structure

Information regarding the functional domains of plant promoters is scarce, and the basis of delineating these domains is mainly dependent on sequence homologies to analogous domains of animal genes. Many plant genes have been sequenced and show evidence of conservation of promoter domains when compared to the animal promoters. Direct testing for the functional importance of such domains has been done to a limited extent by in vitro mutagenesis and introduction of mutated promoters into plants using a variety of methods including: the transformation of plant protoplasts with naked DNA; by natural infection routes either with plant virus such as the brome mosaic

TABLE 1

SUMMARY OF PROMOTER MUTATIONS IN PLANT GENES

GENE	TYPE[a]	PROMOTER MUTATION[b]	PLANT SYSTEM & ASSAY[c]	RESULTS[d]	REFERENCE
1. TDNA OCS	C	5'-deletions: -292 to -116	Nic. tumors: opine assay	-292: +, -168: - -116: -	(82)
NOS	C	5'-deletion: -261	Nic. tumors: opine assay	-261: +	
2. TDNA Tr4	C	Tn5 insertions: -570 to -85	Nic. tumors: rooty phenotype	-570 to -121: + -85: -	(91)
3. TDNA Tr2	C	Tn903 insertion: -72 (from pUC4K)	Nic. and Kal. tumors: shooty phenotype	-72: +	(78)
4. TDNA NOS	C	5'-deletions: -130 to +38	Kal. tumors: NOS assay	-130 to -77: +, -67 to -22: low	(140)
5. TDNA NOS	C	5'- and 3'-deletions: -205 to +78	Nic. tumors: CAT and Nptll assays	3'/ +78 to -17: +, -23 to -40: 12%, -64 to -219: -; 5'/ -205 to -155: +, -130: 36%, -101 to -17: -	(2)
6. CaMV 35S	C	5'-deletions: -343 to -46	Nic. tumors & regen. plants Northern blot analysis Ref. gene: NOS (p)/Nptll	-343; +, -168: 90%, -105: 30%, -46: -	(117)
7. Pea rbcS	I	5'- and internal del.: -1052 to -13	Pet. calli: S1 nuclease and Northern blot analysis Ref. gene: NOS (p)/Nptll	-1052: +, -437: 36%, -352 to -35: 18%, -107/-56: 170%	(109)

TABLE 1 (con't)

GENE	TYPE[a]	PROMOTER MUTATION[b]	PLANT SYSTEM & ASSAY[c]	RESULTS[d]	REFERENCE
8. Pea rbcS	I	5'- and internal del.: -973 to -92	Nic. calli: CAT assay	-973: +, -722: 21%, -357 to -92: 5%, -92/+1: -	(153)
9. Pea Cab	I	5'-deletions: -2500 and -400	Nic. regen. plants, NptII	-2500 to -400: 8-fold decrease	(146)
10. Soybean heat shock	I	5'-deletions: -3250 to -95	Hel. primary tumors: S1 nuclease analysis	-3250 to -1175: +, -95: 5%	(55)
11. Snapdragon chalcone synthase	I	5'- and internal del.: -3900 to -39	Nic. calli and regen. plants, RNA dot blot and Npt II assay, Ref. gene: MnS (p)/NptII	-3900: 80%, -1200: +, -327: 20%, -39: >5%	(75)

[a]"C" refers to constitutive expression and "I" refers to inducible expression.

[b]Deletions are in basepairs from the start of transcription.

[c]Nic. is *Nicotiana tobacum*, Kal. is *Kalanchoe* sp., Pet. is *Petunia hybrida*, and Hel. is *Helianthus anuus*. NptII and CAT assay refers to chimeric promoter/gene fusions either to the NptII or CAT coding sequences and assayed for enzyme activity. Ref. gene refers to the reference gene used as an internal standard; NOS (p)/NptII refers to the nopaline synthase promoter and MnS (p)/NptII refers to the mannopine synthase promoter fused to the NptII coding region.

[d]Percentage values are levels of transcription relative to wild-type, "+" refers to 100%, "-" refers to mutant or 0.1%.

virus (45), or with *Agrobacterium tumefaciens* T–DNA based vectors (65, 114). Table 1 summarizes most of the studies involving plant promoter mutations and the assessment of these gene activities in either callus tissue or regenerated plants. With few apparent exceptions, the overall promoter structure of higher plant genes seems to be similar to that of the typical animal promoter. Therefore, the plant promoter will be discussed within the context of the three major regulatory domains described above for animal genes.

The TATA Box

As with the animal and viral genes, the TATA box seems to be the most conserved sequence motif present in plant promoters (10, 16, 77, 86, 89, 117). By simple inspection, this element can be found approximately 30 bp from the start of transcription and may therefore function in plants in a similar manner to its function in animal genes. Mutations within the TATA element of genes transcribed by plants have been made only in a few cases. An et al. (2) demonstrated the function of TATA in maintaining transcriptional activity by creating 3' deletions into the TATA element of the T–DNA nopaline synthase (NOS) gene from *A. tumefaciens*. The NOS promoter was fused to the CAT coding region, incorporated into the T–DNA, and then introduced into tobacco calli using *A. tumefaciens*. Based on CAT activity assays, the level of expression dropped nearly 10–fold when half of the TATA sequence was removed by a 3' deletion. In another case, Morelli et al. (109) performed 5' deletions on the ribulose bisphosphate–carboxylase gene (*rbc*S) of pea to position –35 bp, keeping the TATA sequence intact, and still retained 18% of wild type activity. When a further deletion was made removing the TATA box, transcriptional activity was not detectable. Based on these limited results, this motif does seem

to be required for efficient expression. However, the other role of positioning the start of transcription has not been directly demonstrated in plants.

The Upstream Element

The most common sequence motif of the upstream element in plants appears to be the CAAT box which is usually positioned between –50 to –120 bp. By examination of a variety of plant promoter sequences, none of the other four consensus sequences present in the upstream element of animal genes (112) are found in analogous positions in plant promoters. A limited number of 5' deletions in the promoter region of the cauliflower mosaic virus (CaMV) 35S gene, as an example, reveal some information concerning the function of the upstream element in plants (117). The TATA box by itself is not enough to produce detectable amounts of transcripts when the 35S gene is deleted to position –41 bp. Inclusion of 5' flanking sequences to –105 bp can, however, maintain approximately 30% of the wild type transcriptional activity. The sequence between –58 and –105 bp of 35S gene promoter that is necessary for this low level expression includes a CAAT box and a 16 bp inverted repeat which shows limited homology with the SV40 enhancer core consensus (51). The CAAT sequences of the *rbc*S gene, however, are not necessary for activity as demonstrated by an internal deletion of this region (109). By removing –56 to –107 bp, *rbc*S transcription actually increases to approximately 170% of normal wild type levels. Another internal deletion made by Timko et al. (153) from –92 to +1 bp of the *rbc*S gene resulted in no activity. Based on these observations the sequences around the CAAT box are not required for efficient light inducible expresson, but the TATA box is essential for such expression. Even though the promoter region of the *rbc*S gene which includes the CAAT box was not needed, this gene seems to require other sequences 5'– to TATA for

activity to occur. This would then reflect the situation seen in some of the inducible animal genes in which short sequence elements are required to be located immediately upstream of the TATA box (121, 135). Direct demonstration in plants of a requirement for the presence of an element immediately proximal to TATA in the region analogous to the upstream element of animal genes is still needed.

Analysis of the 5' deletion mutations in the promoter of the NOS gene demonstrated the requirement of sequences upstream of the TATA box for full transcriptional activity. In one study, Shaw et al. (140) demonstrated that 77 bp of promoter which only includes CAAT box and TATA was required for wild type activity when using *Kalanchöe* tumors and assaying for the production of nopaline. However, when An et al. (2) used tobacco calli instead, to examine deletion mutants of the NOS gene, 155 bp of promoter sequence upstream to the cap site was needed to maintain normal levels of activity. An 8 bp inverted repeat along with the second part of an 11 bp direct repeat reside within the sequences between −101 and −155 bp of this promoter. The presence of repeats in the upstream element region of the NOS promoter is similar to the presence of repeats (different from those found in the NOS promoter) in the promoter of the CaMV 35S gene, and suggests that the redundancy of these sequences may be important for the expression of these two genes.

The opposing results of the two NOS studies may be a preliminary indication of differential transcriptional activity requirements for different portions of the same promoter which is dependent on the host plant species. Another example of differential sequence requirements for expression as a function of the plant system used can be seen with the octopine synthase gene (OCS) as demonstrated by Koncz et al. (82). They made 5'–deletions to –292, –168 and –116 bp of the OCS promoter, introduced them into tobacco tissue and

measured OCS activity. The 5'– deletion to –292 bp did not affect expression, whereas a deletion to either –168 or –116 bp resulted in no activity. This result does not agree with results of Muria et al. (111) who demonstrated that the expression levels of the OCS promoter containing only –116 bp (*Bam*HI) fused to the phaseolin coding region was very high in sunflower plants. Therefore, it would appear that the requirement of sequences upstream of the CAAT box differs depending on the host system.

Enhancer–like Elements

Enhancer–like elements are found in two light–inducible tissue–specific genes of pea, the *rbc*S (43, 153) and the light harvesting chlorophyll a/b binding protein (Cab) (145). In the case of the *rbc*S enhancer–like element, a 240–280 bp region (ca. –330 to –50 bp) imparts light–inducibility as well as tissue–specificity on both the NOS promoter and the CaMV 35S promoter when present 5' to the TATA box in either orientation. In another study, however, a region from –973 to –92 bp of *rbc*S gene, which includes the enhancer–like element failed to promote activity in a downstream location of NOS promoter fused to the CAT coding region (153). The conclusion that the *rbc*S enhancer does not function 3' to the gene may not be valid, since this latter construction might have interfered with poly (A) addition due to the placement of the 881 bp *rbc*S region between the NOS/CAT gene and its poly (A) addition site. It is possible that the *rbc*S enhancer–like region may function normally if it is placed further downstream of the poly (A) addition site. When assayed in regenerated tobacco plants, a 247 bp region (–347 to –100 bp) from the Cab gene also confers light–induction and tissue–specificity on the constitutive NOS promoter when placed upstream in either orientation. Duplicating this enhancer–like element increases light induction 2–fold of the NOS promoter in

leaves when compared to a single element. This element is, however, unable to activate any expression of the NOS/NptII gene while in the downstream position regardless of orientation. Another interesting effect that this 247 bp sequence imparts is its ability to silence, in either orientation, the constitutive expression of the NOS promoter in root tissues. When the enhancer–like sequence is removed, this promoter is equally active both in leaves and roots in its normal constitutive manner. Two different properties are inherent within this Cab enhancer–like element: one involving tissue–specificity, and the other, its ability to increase activity by the introduction of multiple copies of the sequence.

Inducible Plant Genes

Many examples of inducible genes exist in plants (43, 57, 69, 87). Some of these genes contain repetitive elements with enhancer–like characteristics similar to those found in animal genes. Regulation of thermal inducibility has been conserved throughout diverse groups of eucaryotes. As mentioned earlier, heat shock elements (HSE) are small repetitive sequences which interact with specific factors (155) in animal heat shock genes, and are responsible for heat inducibility of transcription. The sequences of the promoters of several heat shock genes from soybean and maize are known (24, 113, 126) and contain 5'–flanking sequences that are similar to the *Drosophila hsp70* gene HSE's. Apparently conservation of other aspects of the thermal induction system have also been maintained between animals and plants as demonstrated by Spena et al. (150). In this study the *hsp70* promoter of *Drosophila* was fused to the NptII coding region and introduced into tobacco tissue using *A. tumefaciens*. The fused gene was expressed after heat shock in 75% of the transformed calli. An example of interspecies conservation of the

heat shock response in plants was illustrated by Gurley et al. (55) in which they introduced a small heat shock gene (*Gmhsp17.5-E*) of soybean into sunflower tumors using a Ti-plasmid based vector system and showed a typical thermal inducible response. Deletions of the 5'-flanking sequences in this gene to -295 bp still retained 70% of its wild type (-1175 bp of promoter) heat inducibility. However, deletions to -90 bp sharply reduced the heat-inducible transcription to approximately 24% of normal levels. There are two regions between -295 and -90 bp that may be responsible for the loss of activity: an imperfect dyad/TATA centered at -250 bp which was found in several *Drosophila* heat shock genes (149), and a 7 out of 10 bp homology with the HSE consensus that is partially disrupted by the -90 bp deletion.

The alcohol dehydrogenase gene of maize, another inducible plant gene, also seems to have a complex promoter structure based on in vivo DNase I hypersensitivity. A region (-160 to -700 bp) in the promoter of the maize alcohol dehydrogenase-1 (*Adh-1*) gene is accessible to DNase I digestion in a constitutive manner whether or not the maize cells are placed under anaerobic stress (122). However, another region from -35 to -150 bp is only accessable to DNase I digestion when anaerobically induced. These DNase I hypersensitivity sites may be due to preferential change in the chromatin structure which is induced by anaerobic conditions to allow access of this region to interaction with factors required for expression.

In another study, Ingersoll, Ferl and Gurley (unpublished results) demonstrated that the mechanism of anaerobic induction of maize *Adh-1* is conserved in heterologous plant species. They introduced the *Adh-1* gene including approximately 1100 bp of 5' flanking sequences into sunflower tumors incited by *A. tumefaciens* and observed expression specifically inducible by anaerobic stress. They also demonstrated that this induction could

still occur, albeit at very low levels, with only –145 bp of 5'–flanking sequence. It seems plausible that a region of the *Adh–1* promoter upstream of position –145 bp may be required for high levels of expression only, and that the TATA–proximal sequences are required for the anaerobic induction. Demonstration of any enhancer–like elements similar to the MRE's or HSE's which can anaerobically induce *Adh–1* gene awaits the results of further studies.

T–DNA of *A. tumefaciens*

The T–DNA which resides on the tumor inducing plasmid (pTi) of *A. tumefaciens*, is transferred and randomly integrated into the plant genome upon wounding. Expression of phytohormone–producing T–DNA genes then induces crown gall formation (for review see 114). The opine synthases are also encoded on the T–DNA, and are involved in the production of opines (usually condensation products of certain amino acids and sugars) which supply the *Agrobacterium* with a carbon and nitrogen source (123). The strains of *A. tumefaciens* and the tumors they can incite are categorized according to the type opine (for example, octopine, nopaline, agropine, ect.) produced in the tumor tissue (22).

In plant tumors incited by some octopine producing strains of *A. tumefaciens*, the T–DNA is separated into two parts, T–left and T–right (Fig. 1–1). T–left encodes phytohormone biosynthetic genes as well as the octopine synthase (OCS) gene, while T–right only seems to encode opine synthases (37, 118). Three genes are responsible for phytohormone production: 1, 2, and 4. Genes 1 and 2 are involved with auxin biosynthesis, while gene 4 is involved with cytokinin biosynthesis (14, 157). All three are responsible for tumor formation. Gene 2 which encodes a tryptophan–2–monooxygenase shows

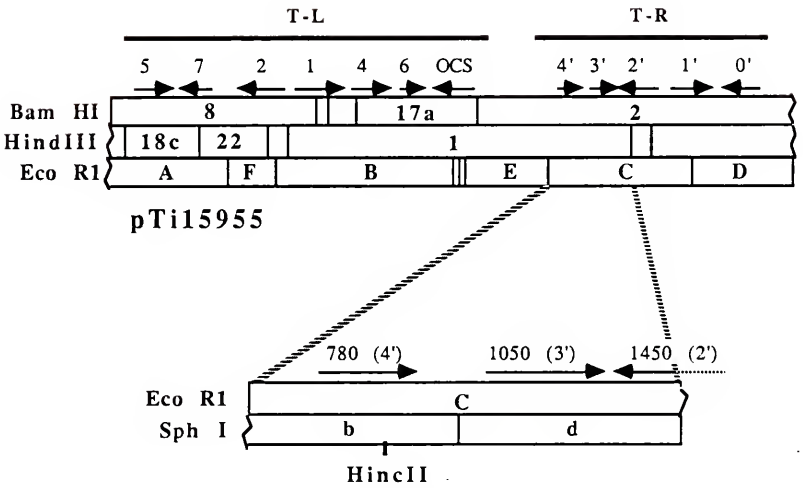


Figure 1–1. T–DNA restriction enzyme map of pTi15955. The restriction sites are based on the sequence of Barker et al. (5). Arrows denote the position and polarities of the transcripts present in T–left (165) and T–right (128). The *Eco*R1–*Sph*I fragment containing the 780 gene was used in this study. Designations 1050 and 1450 correspond to transcripts 3' and 2' in T–right.

significant sequence homology with the indoleacetamide hydrolase gene (iaaM) of *Pseudomonas savastanoi* (173). Amino acid residues 239–263 of the gene 1 protein show a high degree of homology to the adenine binding site (amino acid residues 5–29) of the *Pseudomonas fluorescens* p–hydroxybenzoate hydroxylase. The homology with genes from *Pseudomonas*, and the fact that some of the T–DNA genes are also expressed in *A. tumefaciens* (114) suggest that T–DNA genes may have an ancestrally bacterial origin.

The complete sequence of the T–DNA from an octopine–type pTi15955 strain is known (5), and indicates 26 open reading frames of longer than 100 amino acids in the T–left and T–right. Only 12 of these open reading frames correspond to those genes known to be transcribed in plants (56, 110, 167). Most of these plant–transcribed genes have characteristics of eucaryotic genes transcribed by RNA polymerase II. These include the TATA sequences and CAAT box motifs upstream from the start of transcription and poly (A) addition signals beyond the stop codon. There is no evidence that any T–DNA genes contain introns (110, 166, 167).

The genes present on the T–DNA are transcribed in plants in moderate to low abundance representing less than 0.001% of the total poly (A) RNA of the tumor cells (166). T–DNA genes are usually thought to be expressed in a constitutive manner regardless of the plant tissue (11, 67, 120). However, gene 5, the left–most gene on T–left, may be expressed in a tissue–specific manner (81) even though the results of the gene 5 studies are unclear. The opine levels from the gene 5 promoter/OCS coding region fusion gene appears to be expressed at the highest levels in callus tissues and in stems of transformed tobacco plants and are barely detectable in fully developed leaves. The expression of this fusion gene in matured leaves is fully restored when leaf sections are incubated on a high auxin, low cytokinin medium. On the other

hand incubation of callus tissue on high cytokinin, low auxin medium results in a decrease in activity of this gene. These results suggest that the product of gene 5 is produced only in tissues having a high level of internal auxins relative to the level of internal cytokinins (81). Alternatively, since transcript levels were not directly assessed, these results may only reflect the differential availability of opine precursors in the various tissues after various hormone treatments.

Agrobacterium has a very broad host range with tumor formation observed on at least 643 host plant species including 310 genera of dicotyledenous plants, 4 families of monocots and 43 species of gymnosperms (26, 64, 136). In order to incite tumors, the T–DNA genes must be expressed to some degree and therefore the promoter sequences of these genes must contain elements which are conserved throughout most of the plant kingdom.

A T–right gene from pTi15955 used in this dissertation was designated as the 780 gene since the corresponding transcript is approximately 780 bases in length (73, 168). Previous studies on the 780 gene include an approximate determination of the start of transcription (168) which placed it approximately 30 bp upstream of the first methionine residue. The abundance of the 780 gene transcripts in poly (A) RNA from the tobacco tumor line E1, as analyzed on northern blots, was much higher than that from poly (A) RNA from the sunflower tumor line, PSCG–15955. This differential abundance is probably due to the greater copy number of T–right present in the genome of E1 tumor tissue. Karcher et al. (73) reported that this gene was trancribed about as equally well in both tobacco (E9) and sunflower (S4–2) tumor tissue culture lines at levels much higher than the other T–right genes. They observed differences of the relative level of 780 transcripts in other tobacco and sunflower calli. The 780 gene, therefore, seems to be expressed at fairly high

levels in at least two plant species. Due to the lack of definitive information to the contrary, the 780 gene promoter is assumed to be constitutive in its expression.

The function of the 780 gene as yet is still unknown. Transposon Tn5 insertions in four locations within and upstream from the 780 gene do not affect the production of agropine and mannopine in callus tissue (128). The oncogenic properties of the octopine Ti plasmids and the transfer and integration of T–left and T–right are unaffected by mutations in any of the T–right genes. The same T–right mutations also do not affect excretion of mannopine or agropine since opines are detected in the media of the tumors involved including a tumor line in which T–left is deleted (128).

Since little is known concerning constitutive plant promoters, a characterization of the functional domains of the 780 gene promoter is of particular importance. The compact nature of the T–DNA genes and the availability of the 780 gene make it amenable to promoter studies in plants. In order to identify DNA sequences involved in the transcriptional expression of the 780 gene, 5' and internal deletions in the promoter were made, and then introduced into sunflower seedlings using a Ti plasmid vector system developed for this study. The level of transcription for each promoter mutation was accurately determined using a homologous reference gene as an internal standard which is similar in principle to that developed by McKnight and Kingsbury in the analysis of linker scan mutations of HSV tk gene (103). From this analysis, three functional domains within the 780 promoter were identified. One of these promoter domains seemed to have some of the characteristics of an enhancer and was designated as an activator element. The activator was further analysed for enhancer–like characteristics by testing fusions with the 780 gene TATA on a reference gene vector. A final

study was also made to detect interactions of transcriptional factors present in crude nuclear extracts with the activator.

CHAPTER 2

PROMOTER MUTATION ANALYSES

Introduction

A T–DNA gene was used as a model for identifying sequences required for constitutive transcriptional control in higher plants. The effects of modified 5'–flanking regions of the 780 gene on transcription was assessed in vivo using the natural transformation scheme of *A. tumefaciens*. A Ti plasmid based vector system was first developed in order to transfer this gene into the genome of a host plant. This vector system involved the construction of a double gene shuttle vector which contains two copies of the 780 gene. One of these copies was a mutated test gene while the other acted as reference; both integrated into the T–left DNA of a strain of *A. tumefaciens* 15955 which had its endogenous 780 gene deleted. The start of transcription was then accurately determined to the nucleotide. The test gene could also be distinguished from the reference gene by S1 nuclease mapping using a single end–labeled DNA probe.

The 5'–flanking region of the 780 gene was altered by a series of 5' and internal deletions or substitutions. The effect of these mutations on transcriptional activity in sunflower tumors was assessed by S1 nuclease protection analysis directly utilizing the homologous reference gene as an internal standard. Results from assays conducted at different times were

comparable, since both test and reference transcripts could be detected using a single end-labeled DNA probe. In other plant promoter mutation studies including those using the CaMV 35S gene (117), the *rbc*S gene (109), and the chalcone synthase of *A. majus* (75), heterologous reference genes were used to directly assess transcriptional activity (see Table 1-1). However, the present study was the first to utilize a homologous gene for an internal standard eliminating any possible discrepencies due to differences of test and reference promoters.

Materials and Methods

Removal of the 780 Gene from pTi15955

The endogenous 780 gene of *A. tumefaciens* 15955 (str^r) was removed from T-right in order to facilitate the analysis of mutated 780 gene promoter constructions introduced into T-left. The T-right deletion mutant of *A. tumefaciens* was designated as Ag5260. The 4.7 kbp *Xho*I-*Hin*dIII fragment (15,208 to 19,953 bp, [5]) containing the 780 gene (ORF 18), the 1050 (ORF 21), and most of the 3' end of the 1450 gene (ORF 24) (5, 168) was replaced with a 1.5 kbp *Xho*I-*Hin*dIII fragment from the transposon Tn5 (127) which confers kanamycin resistance (Fig. 2-1). The substitution of the T-right fragment was achieved by the double homologous recombination procedure described by Matzke and Chilton with some modifications (98).

This double homologous recombination event utilized a broad host range vector containing T-DNA fragments and the Tn5 fragment, and was constructed as follows. The *Bam*HI-*Xho*I (13,774 to 15,208 bp, [5]) and *Hin*dIII-*Eco*R1 (19,953 to 21,631 bp, [5]) fragments from T-right that flank the 4.7 kbp *Xho*I-*Hin*dIII region in T-right were ligated into the broad host range

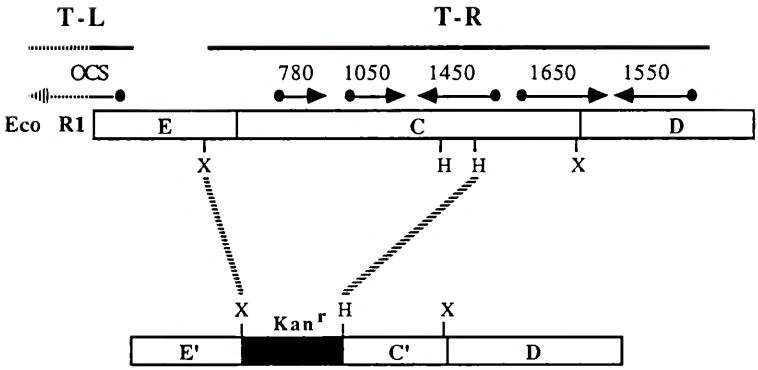


Figure 2–1. The removal of the 780 gene from the T–DNA of pTi15955. An *Eco*R1 map of T–right from pTi15955 is shown at top with the arrows denoting the positions and directions of the transcripts (73, 168). The natural copy of the 780 gene was removed by replacing the 4.7 kbp *Xho*I–*Hin*dIII fragment from the T–right DNA with a 1.5 kbp *Xho*I–*Hin*dIII kanamycin resistant fragment (solid black) from the transposon Tn5 as described in Materials and Methods. Kan^r, kanamycin resistant gene; X, *Xho*I; H, *Hin*dIII.

plasmid pRK290 (31). The 1.5 *Sal*I–*Hin*dIII Tn5 fragment was then inserted between the two fragments resulting in plasmid pKn306 (Fig. 2–2). *Escherichia coli* LE392 harboring pKn306 was grown overnight at 37° C in Luria broth (LB) containing tetracycline (tet) (12.5 μg/ml) and kanamycin (kan) (50 μg/ml). The overnight culture (1 ml) was pelleted in a 1.5 ml microfuge tube. The cells were washed with fresh LB to remove antibiotics and finally resuspended in 1 ml of LB. A portion of this suspension (200 μl) was mixed with an equal volume of similarily treated *E. coli* LE392 harboring pRK2013 that was grown overnight at 28°C in LB containing kanamycin (50 μg/ml). The recipient, *A. tumefaciens* 15955 (str^r) was also grown overnight in LB containing streptomycin (250 μg/ml), washed with 1 ml of LB and 200 μl of this overnight growth was mixed with the two *E. coli* culture mixture. The plasmid pRK2013 encodes conjugal transfer functions which facilitate the tranfer of pKn306 to *A. tumefaciens*. The bacterial suspension was spotted on an LB plate and incubated for 48 hours at 28°C. The cells were resuspended in approximately 2 ml of sterile 10 mM $MgSO_4$, and 200 μl of this suspension was mixed with 300 μl of an overnight growth of *E. coli* SK1590 harboring the plasmid pPH1JI (gen^r) (8). This mixture was spotted on another LB plate and incubated at 28°C overnight. The pPH1JI plasmid is incompatible with pKn306 and its presence allowed selection for the loss of the plasmid pKn306 after recombination with the Ti plasmid. After resuspension of the conjugated cells in 2 ml of sterile 10 mM $MgSO_4$, they were plated on AB minimal media (18) containing streptomycin (250 μg/ml), kanamycin (20 μg/ml) and gentamicin (100 μg/ml). Individual colonies were picked after incubating 3 days on selection plates at 28°C, screened with the 3–ketolactose test for the presence of *A.tumefaciens* (9) and then rescreened for sensitivity to tetracycline at 5 μg/ml. *A.tumefaciens* colonies that were str^r, kan^r, and tet^s should contain the kan^r gene present on

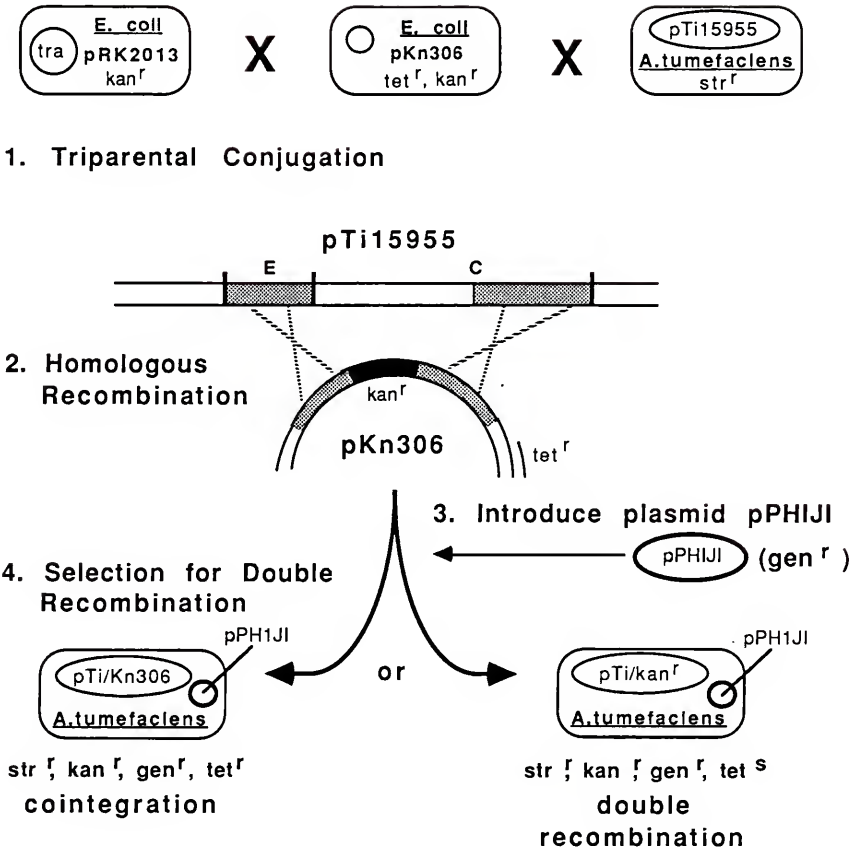


Figure 2–2. Double homologous recombination event. (**1**.) A triparental mating involving two *E. coli* hosts harboring the plasmids pKn306 and pRK2013 separately and the *A. tumefaciens* strain 15955 results in a transfer of the pKn306 into the *A. tumefaciens*. (**2**.) The homologous recombination occurred between pTi15955 and pKn306. (**3**.) After selecting for streptomycin (str^r), kanamycin (kan^r), and tetracycline (tet^r) resistances, the plasmid pPHIJI (gen^r)was then introduced into the *A. tumefaciens* containing pKn306. (**4**.) Subsequent selection for streptomycin, kanamycin and gentamicin resistances as well as tetracycline sensitivity (tet^s) distinguishes colonies that resulted in the double recombination from those that the pKn306 just cointegrated into the Ti plasmid.

the Ti plasmid with subsequent loss of the remainder of the pKn306 plasmid (loss of tet^r). Four tet^S colonies were picked and grown overnight in LB with streptomycin (250 μg/ml) and kanamycin (20 μg/ml) for subsequent analysis.

The Ti plasmid from these four transconjugants was isolated using a small scale DNA preparation described by Casse et al. (15). The resultant DNA was separated on a 0.7% agarose gel which was then treated with a 2% HCl solution for 15 minutes at room temperature. The gel was washed with distillled H_2O, and the DNA was then denatured and transferred to a nitrocellulose filter paper as described by Southern (148). Two different nick translation probes were prepared (95): one from the *Eco*R1–*Hin*dIII fragment containing the 780 gene, and the other from the plasmid (pKS–4, gift from D. Sutton) that contains the Tn5 *Sal*I–*Hin*dIII fragment conferring kanamycin resistance. Both of these probes were hybridized as previously described (55) to duplicate Southern blots to confirm the double homologous recombination event.

The Ti plasmid resulting from this recombinational event lacks the left border sequence of T–right, the genes corresponding to the 780 and 1050 transcripts, and most of the 3' terminus of the gene encoding the 1450 transcript. Plasmid pPH1JI, residing within Ag5260, seemed to interfere with the introduction of shuttle vectors, and was therefore removed by a carbinicillin–cycloserine enrichment procedure (107). This mutant, *A. tumefaciens* Ag5260, was used as the recipient for shuttle vectors containing various constructions of the 780 gene.

Construction of the Intermediate Shuttle Vector Containing the 780 Reference Gene

The 780 gene was initially subcloned from plasmid p403 (56, 110) by ligation of the isolated *Eco*R1–*Sph*I fragment (16,202 to 17,601 bp, [5]) into

pUC–19 (116). The resultant plasmid, designated pUC–19:780, contained the complete gene consisting of approximately 200 bp downstream of the poly (A) addition signal, the protein coding region (138 amino acids), and 476 bp of 5'–flanking sequences.

For precise quantitation of the activity of the promoter mutants, a reference gene was constructed and cloned into the shuttle vector, pW9, as shown in Fig. 2–3. The plasmid pW9 contained a 4.2 kbp *Bam*HI–*Sph*I fragment of T–left (pTi15955) from p233G (55) inserted into the *E. coli* plasmid, pACYC184 (16). The T–left fragment provided the site for homologous recombination of the shuttle vector into the Ti plasmid of *A. tumefaciens* Ag5260. The reference gene consisted of the 780 gene with 290 bp of 5'–flanking sequences and an 8 bp internal deletion in the untranslated leader sequence resulting from the removal of a *Taq*I fragment. The –290 bp 5'–deletion mutant was used to eliminate other upstream *Taq*I sites that would interfere with the construction of the reference gene. The intermediate shuttle vector, designated as pW9–TD, was completed by cloning the *Sal*I–*Sph*I fragment containing the reference gene into pW9.

Promoter Deletion and Duplication Mutants.

A series of 5'– and 3'–deletions covering the 5'–flanking region were obtained as outlined in Fig. 2–4. For 3'–deletions the *Eco*R1 site was changed to *Bam*HI by linker addition (95). Plasmid pUC–19:780 DNA (5 µg) was linearized either by *Eco*R1 (for 5'–deletions) or *Hin*cII (for 3'–deletions) and subsequently digested with the exonuclease Bal 31 (Bethesda Research Laboratories) in a 50 µl reaction at 50 units/ml. The reactions were terminated by adding 1/10 volume of 0.2 M [ethylene–bis(oxyethylene–nitrile)] tetraacetic acid (EGTA, Sigma). Following addition of *Sal*I linkers (95) the 5'– and 3'–deletion fragments

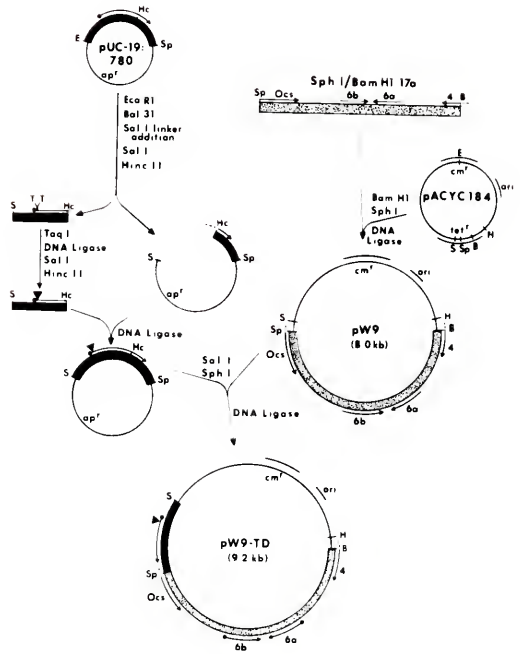


Figure 2–3. Construction of the reference gene and pW9–TD. The reference gene (solid black) was formed by removing an 8 bp *Taq*I fragment (inverted triangle) in the 5' leader region of the 780 gene. The intermediate shuttle vector, pW9–TD, was obtained by ligation of the reference gene into pW9 which consisted of a *Sph*I–*Bam*HI fragment (*Bam*HI fragment 17 [5]) from T–left inserted into pACYC184 (16). B, *Bam*H1; E, *Eco*R1, H, *Hin*dIII; Hc, *Hin*cII; S, *Sal*I; Sp, *Sph*I; T, *Taq*I. ap^r, cm^r, and tet^r are ampicillin, chloramphenicol, and tetracycline resistance, respectively. OCS is octopine synthase transcript. 4, 6a and 6b are T–left transcripts of the T–DNA (165).

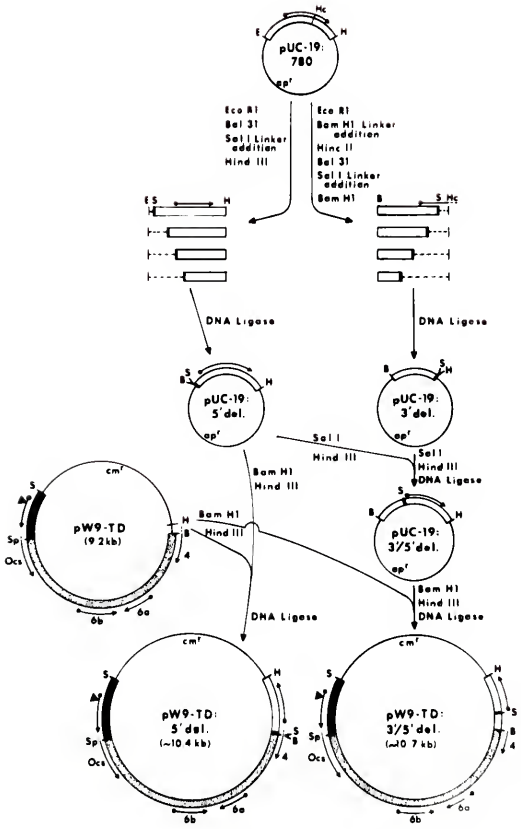


Figure 2–4. Construction of the double gene shuttle vector containing the 5' and internal deletions of the 780 gene. 5' and 3' deletions were constructed by Bal 31 digestion of the 5' flanking region and subsequent recloning of the deletion mutants into pUC–19. The 5' deletion clones were then ligated into pW9–TD, generating double gene shuttle vectors. Internal deletions were made by ligating isolated *Sal*I–*Hin*dIII 5' deletion fragments into the 3' deletion clones as decribed in Materials and Methods. B, *Bam*H1; E, *Eco*R1, H, *Hin*dIII; Hc, *Hin*cII; S, *Sal*I; Sp, *Sph*I. ap^r and cm^r are ampicillin and chloramphenicol resistance, respectively. The black triangle denotes the *Taq*I deletion in the reference gene leader. OCS is octopine synthase transcript. 4, 6a and 6b are T–left transcripts of the T–DNA (165).

were separated from pUC–19 sequences by digestion with *Hin*dIII and *Bam*HI respectively, and cloned into either the *Sal*I–*Hin*dIII or *Sal*I–*Bam*HI sites of pUC–19. Deletion endpoints were determined by DNA sequencing (100, 130). Internal deletion and duplication mutants were constructed after selection of appropriate pairs of 5'– and 3'–deletion clones. The appropriate 5'–deletion fragments (*Sal*I–*Hin*dIII) were isolated and then inserted into the *Sal*I–*Hin*dIII site of the matching pUC–19:3'–deletion clones. The 5'– and internally deleted 780 genes were finally cloned between the *Bam*HI–*Hin*dIII sites of pW9–TD to form double gene shuttle vectors.

Triparental Conjugation and Tumor Formation

Double gene shuttle vectors carrying the various 780 gene promoter alterations were transferred from *E. coli* LE392 into *A. tumefaciens* Ag5260 by the triparental conjugation procedure of Fraley et al. (44) described as follows. *E. coli* LE392 harboring a double gene shuttle vector (cm^r) and another *E. coli* LE392 harboring the transfer–helper plasmid, pRK2013 (kan^r) encoding the conjugal transfer functions, were separately grown overnight at 37°C in LB, with appropiate antibiotics. *A. tumefaciens* Ag5260 was also grown for two days at 28°C in LB with appropriate antibiotics. After washing 1 ml of each of these cultures with fresh LB, 200 µl of each were mixed together and spotted on a LB plate and incubated 2 days at 28°C. The cells were then resuspended in 2 ml of sterile 10 mM $MgSO_4$ and plated on AB minimal media with streptomycin (250 µg/ml), kanamycin (20 µg/ml) and chloramphenicol (15 µg/ml) and incubated at 28°C for 3–5 days. Transconjugant colonies were tested by overnight growth in 2 ml of LB with streptomycin (250 µg/ml), kanamycin (20 µg/ml) and chloramphenicol (2.5 µg/ml). The str^r, kan^r, and cm^r clones were analysed by Southern blot analysis (148) after small scale DNA preparation (15).

Tumors were incited on one week old sunflower seedlings (*Helianthus annuus* cv. Large Grey) by injecting a drop of an overnight growth of *A. tumefacien* transconjugant into the hypocotyls using a syringe fitted with a 25 gauge needle (5/8 inch). The plants were then grown with an 18 hour light cycle for 14 to 16 days at room temperature. Usually 200 to 300 tumors for each promoter mutation were harvested, immediately frozen with liquid nitrogen, and then stored at –70°C.

Total RNA Extraction and Poly (A) RNA Isolation from Sunflower Tumors

Total RNA was extracted by the method of Jackson and Ingle (68) as a modified by Czarnecka et al. (23, 55). Approximately 50 g of frozen tumors were ground in a mortar and pestle in the presence of 200 ml of grinding buffer containing 100 mM Tris–HCl (pH 8.8), 500 mM NaCl, 6% (wt/v) p–aminosalicylate (Sigma), 2% (wt/v) triisopropylnapthalenesulfonate (Kodak), 6% (v/v) n–butanol, 0.2% (wt/v) polyvinylpyrrolidone (Sigma), 0.3 mM diethyldithiocarbamate (Sigma), and 0.13 mM dithiothreitol (DTT, Sigma). After grinding the tumors to a slurry, the mixture was completely homogenized for 2–3 minutes using a Tekmar tissumizer, then extracted twice by shaking for 1 hour with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The nucleic acids were precipitated by adding 1/9 volume of 3 M sodium acetate and 2 volumes of ice–cold 95% ethanol.

Following centrifugation of the precipitate, the collected pellet was resuspended in 20 ml of 10 mM Tris–HCl (pH 8.0), 1 mM EDTA and 1% sarkosyl (Sigma) (TES) and extracted again with phenol/chloroform for 30 minutes before centrifuging. An equal volume of a precipitating solution containing 4 M LiCl, 4 M urea and 2 mM EDTA was added to the separated aqueous phase to cause the RNA to precipitate during an overnight incubation at 4°C. The RNA

was collected by a 20 minute 6000 Xg centrifugation and washed with the original volume of a 1/2 strength precipitating solution, then the RNA was collected again by centrifugation. The pellet was resuspended in 5 ml of TES and precipitated with sodium acetate and 95% ethanol. The total RNA was finally resuspended in sterile 10 mM Tris–HCl (pH 8.0) and 1 mM EDTA to a concentration of 15 mg/ml.

The poly (A) RNA fraction was isolated using oligo (dT) cellulose chromatography (144). Approximately 5 µg of total RNA (0.5 mg/ml) in 10 mM Tris–HCl (pH7.5), 1 mM EDTA (TE) with 0.1% sodium doedecylsulfate (SDS) was heat denatured (10 minutes at –70°C) and 1/10 volume 5 M NaCl was added, and then the mixture was cooled to room temperature. This RNA solution was cycled 3 times at room temperature through a 1–2 ml oligo (dT) cellulose (Bethesda Research Laboratories) column that was previously equilibrated with TE containing 0.5 M NaCl. After washing the column with the equilibrating buffer, the poly (A) fraction was eluted with TE at 68°C. Generally, the first 2 ml of eluate were pooled and the poly (A) RNA precipitated with both 1/10 volume of 100 mM Mg acetate, 3 M Na acetate, and 2.2x volume of 95% ethanol. The poly (A) RNA was then dissolved in sterile water. Concentrations of RNA were calculated from the optical density at 260 nm (40 µg/ml = 1 O.D.).

S1 Nuclease Analysis

Transcript levels of both the 780 test gene and reference gene were assayed by S1 nuclease hybrid protection (55). Analyses were performed with approximately 15 µg of poly (A) enriched RNA. The hybridization probe was isolated from the 5'–deletion clone pΔ–74 and was 5'–end labeled at the *Hpa*II site located at +60 bp. This probe (*Sal*I–*Hpa*II fragment) contains the wild type leader and 74 bp of 5'–flanking sequences of the test gene. Poly (A) RNA was

hybridized with an excess of the doubled stranded DNA probe (40,000–50,000 cpm) overnight at 38°C. After S1 nuclease digestion (50 units/ml) at 23°C for 30 minutes, the protected hybrids were fractionated on an 8% sequencing polyacrylamide gel containing 7 M urea and exposed (with intensifier screens) to XAR–5 (Kodak) film for 1 to 2 days at –70°C.

Relative transcript levels (RTL) of each mutant were determined by cutting portions of the gel corresponding to radioactive bands and using the gel portions for Cerenkov counting. The RTL values were defined as the ratio of cpm values of the test gene transcripts to cpm values of the reference gene transcripts divided by the ratio of cpm values of the test gene transcripts to its reference gene transcripts (30). All RTL values represent an average of 3 or more independent hybridization experiments.

Primer Extension Analysis

Primer extension analysis was performed to confirm the S1 nuclease analysis and to determine if the reference gene transcripts could be distinguished from wild type gene transcripts. Poly (A) RNA (approximately 15 μg) was added to 0.003 pmoles of a 5'–end labeled primer that was complementary to the region from +33 to +49 bp of the 780 gene mRNA transcript. A final volume of 10 μl of 10 mM Tris–HCl (pH 7.0) and 1.0 mM EDTA was sealed in a 20 μl capillary tube, and placed in a boiling water bath. After 3 minutes, the capillary tube was then quickly transferred to 50°C for 6 hours to allow the primer to anneal to the RNA. After annealing, the contents of the capillary tubes were added to 10 μl of a primer extension buffer (80 mM Tris–HCl [pH 7.0], 150 mM KCl, 20 mM DTT, 6 mM $MgCl_2$ and 1 mM each of dATP, dGTP, dCTP, and dTTP) with 200 units of murine maloney virus reverse transcriptase (Bethesda Research Laboratories) and incubated at 37°C for 30

minutes. The reaction mixture was then phenol/choloroform extracted, precipitated with ethanol and finally fractionated on 8% polyacrylamide gel containing 7 M urea.

Results

Southern Blot Analysis of the T-right DNA of *A. tumefaciens* strain Ag5260

The substitution of the 4.7 kbp region of T-right of pTi115955 for the 1.5 kbp Tn5 fragment conferring kanamycin resistance was examined by Southern blot analysis. The Ti-plasmid bands from wild type *A. tumefaciens* strain 15955 and an intermediate transconjugant containing the plasmid pKn306 (introduced without selection) hybridized as expected to the 780 gene probe. The Ti-plasmid of four clones after selection for recombination (lanes 2-5, Fig. 2-5) did not hybridize to the 780 gene probe but instead hybridized to the kanamycin gene probe which has no homology to the wild type Ti-plasmid. This clearly demonstrates that the endogenous 780 gene of pTi15955 has been replaced with the kanamycin gene in strain Ag5260 (lane 2, Fig. 2-5).

Triparental Conjugations with Double Gene Shuttle Vectors

Double gene vectors containing the reference gene alone or the test and reference gene together, were transferred into Ag5260 using the triparental conjugation method. Between 5-50 colonies arose after 4-5 days on AB minimal plates with streptomycin (250 µg/ml), kanamycin (20 µg/ml), and chloramphenicol (15 µg/ml). Of these potential transconjugants approximately 50% grew overnight in 2 ml LB in the presence of the same level of streptomycin and kanamycin as was present in the AB minimal plates but with reduced chloramphenicol (5-2.5 µg/ml). These clones were then used to

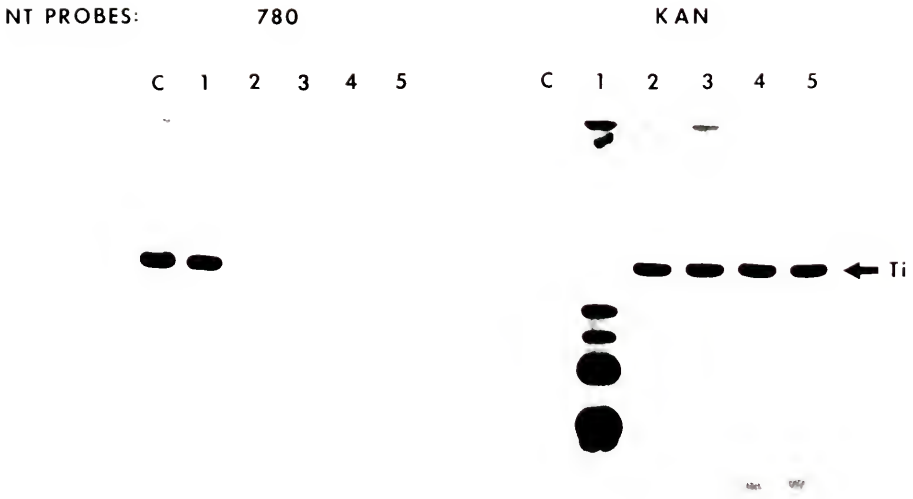


Figure 2–5. Southern blot analysis of the double homologous recombination event. Approximately 1–2 μg of total DNA isolated from *A. tumefaciens* clones was used for each lane. Nick translation probes used were the isolated *Eco*R1–*Sph*I fragment containing the 780 gene and the isolated *Xho*I–*Hin*dIII fragment containing the kanamycin phosphotransferase gene from Tn5. Lane C contains DNA from strain 15955. Lane 1 contains DNA from an *A. tumefaciens* transconjugant clone harboring pKn306 plasmid. Lane 2–5 contain DNA from four separate transconjugants after selection for the double recombination. Ti denotes the position of the Ti plasmid.

inoculate sunflower plants. The 780 gene used as a nick-translated probe, hybridized to the Ti-plasmid bands of the various clones when analyzed by Southern blotting (Fig. 2-6) as well as to the positive control, wild type strain 15955. These results along with the lack of hybridization of the recipient, Ag5260, confirm the transfer and integration of the shuttle vector into the Ti plasmid.

Discrimination of the 780 Test and Reference Gene Transcripts

S1 nuclease hybrid protection mapping was performed to accurately determine the 5' start of transcription and to quantitate the transcripts from the 780 test and reference gene. The deletion of the *Taq*I fragment in the untranslated leader of the reference gene resulted in shorter transcripts that could be distinguished from the longer test gene transcripts. However, the basis of transcript discrimination obtained using S1 nuclease assay was due to a local region of nonhomology between the wild type DNA probe and the reference gene transcript which allowed the S1 nuclease to cleave the resulting 8 bp loop in the labeled DNA (Fig. 2-7). As illustrated in Fig. 2-7, protection of the labelled probe by the reference gene RNA resulted in a smaller fragment than protection by the test gene RNA.

The autoradiographs of S1 nuclease mapping gels presented in Figs. 2-8 (and 2-10) show the length of protected hybrids obtained using the test gene DNA probe with RNA derived from the wild type gene, the reference gene, and with RNA from the test and reference genes present in the same vector. The major start site of transcription for the wild type gene maps 60 bp upstream of the *Hpa*II site. When RNA from the tumors containing the reference gene was hybridized to the test gene DNA probe, a cluster of bands ranging in size from 46 bp to 54 bp was observed. The position of this cluster of protected bands

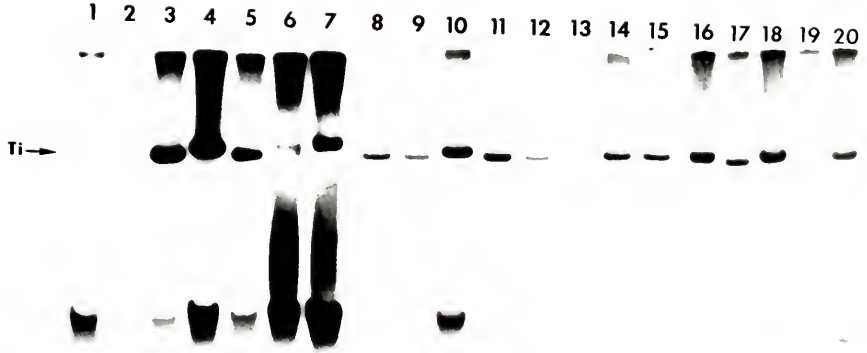


Figure 2–6. Southern blot analysis of selected double gene transconjugants. Lane 1 contains approximately 200 ng of DNA from *A. tumefaciens* strain 15955. Lane 2 contains approximately 1–2 µg of total DNA isolated from *A. tumefaciens* strain Ag5260. Lanes 3–20 contain approximately 1–2 µg of total DNA isolated from various *A. tumefaciens* strain Ag5260 clones containing double gene shuttle vectors recombined into the Ti plasmid. The nick translated probe is the *Eco*R1–*Sph*I fragment containing the 780 gene. Ti denotes the position of the Ti plasmid. The lower bands are the chromosomal DNA.

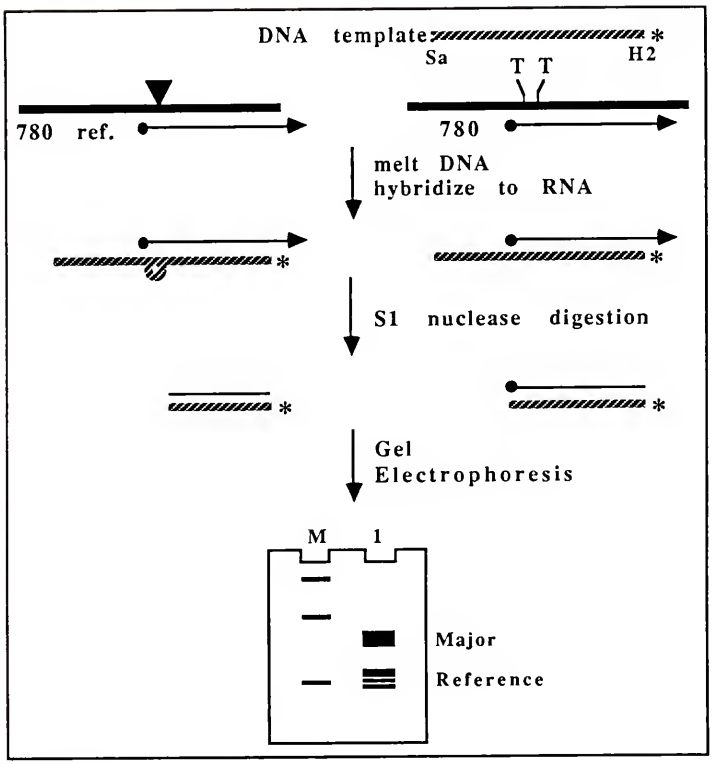


Figure 2–7. S1 nuclease hybrid protection strategy. The poly (A) RNA from sunflower tumors containing both the test and reference gene was hybridized to the 5' end–labeled DNA probe from the wild type 780 gene. Upon forming a hybrid with the DNA template, the transcript from the reference gene formed a single–stranded loop due to its deletion whereas the transcript from the test gene protects the DNA probe in a normal fashion. After S1 nuclease digestion, two sizes of protected DNA resulted which were then separated by electrophoresis on a denaturing gel. The hatched line represents the DNA probe (–74 to +60 bp) with the asterisk denoting the site of the 5' label. The inverted triangle represents the *Taq*I deletion in the 780 reference gene leader sequence. Sa,. *Sal*I; H2, *Hpa*II; T, *Taq*I; M, marker lane; 1, lane containing the protected end-labeled fragments.

Figure 2–8. S1 nuclease mapping of the 5' termini of poly (A) RNA homologous to the 780 gene. A DNA sequence ladder was utilized for sizing S1 nuclease hybrid–protected fragments from wild type (3 μg of RNA) and the reference genes (15 μg of RNA). The M lane contains ΦX174 DNA *Hae*III–generated marker. The arrows denote 5' termini of test gene transcripts and the 8 bp loop of the probe due to the *Taq*I deletion in the reference gene transcripts. The sequence at the right of the ladder represents the coding strand which is the complement of the actual DNA ladder sequence.

corresponded to the predicted position of the 8 bp loop in the probe:RNA hybrid. When RNA from tumors containing both test and reference genes was analysed (lanes WT and Ref., Fig. 2–10), two clusters of protected hybrids were seen at the predicted positions, demonstrating the ability of this method to assess the relative abundance of transcripts derived from these two promoters. Primer extension analysis also confirmed that the 5' termini of the reference gene is 8 bp shorter than the test gene (Fig. 2–9) and that the two types of transcripts can be distinguished from each other.

A weakly protected band of 120 bp was also observed which suggested the presence of another 5'–terminus 60 bp upstream from the major start site. The level of transcripts originating from this minor start site relative to the major start site appeared to be 10% of the major transcript level as determined by the S1 nuclease mapping procedure. The true abundance of the minor transcript, however, may be only 1–2% of the major transcript based on the results of primer extension analysis (data not shown). With the DNA probe in excess, the longer transcripts from the minor promoter seemed to be more stable than the transcripts from the major promoter and reference gene while using S1 nuclease mapping conditions. When using primer extension analysis the primer had no preference for any of these three transcripts and hybridized with them equally well. This allowed a better estimation of the minor transcripts relative to the major transcripts. Other bands were also seen in the primer extension analysis which did not correspond to the S1 nuclease hybrid protected bands. These minor bands were probably due to the primer hybridizing to other RNA molecules since these bands could be seen in the primer extension analysis when RNA from sunflower tumors not containing the 780 gene was used (Fig. 2–9).

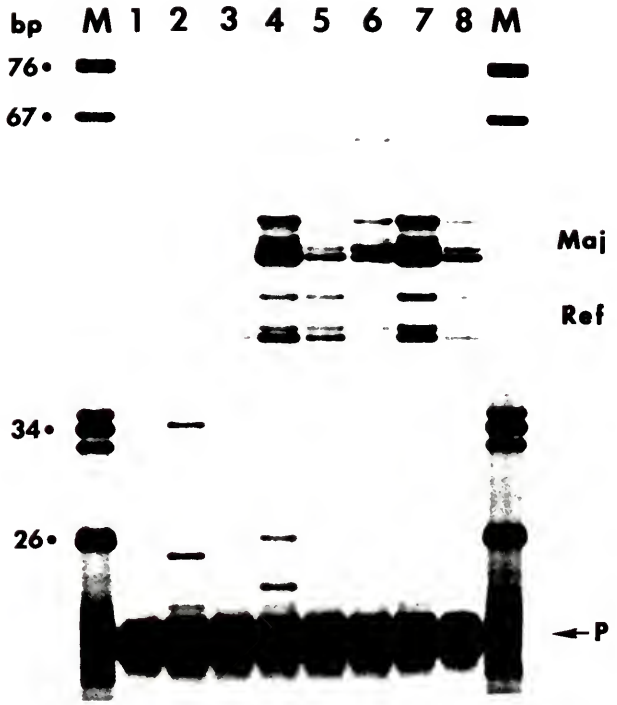


Figure 2–9. Primer extension analysis of RNA derived from 780 test and reference genes. Each lane (except for Lane 1 and 6) contains approximately 15 µg poly (A) RNA isolated from sunflower tumors incited with various *A. tumefaciens* clones and hybridized to a 5' ^{32}P–end–labeled primer (17 nucleotides long from +33 to +49 bp of the 780 leader region). This hybrid was elongated with reverse transcriptase as described in the Materials and Methods. Lane **1** contains 15 µg of yeast tRNA. Lanes **2–8** contain tumor RNA incited by *A.tumefaciens* harboring the following 780 constructions: **2**, Ag5260; **3**, reference gene alone; **4**, wild type 780 gene and reference gene; **5**, Δ–290; **6**, 5 µg of poly (A) RNA from tumors with the wild type 780 gene and reference gene; **7**, ID–112/–74; **8**, ID–76/–74. Maj, major transcript; Ref, reference transcript; P, end–labeled primer. Lane M contains pBR322 *Hpa*II–generated DNA marker.

Figure 2-10. Autoradiograph of S1 nuclease analysis of 780 gene mutants. Approximately 15 µg of poly (A) RNA was used for all lanes. Lanes WT, REF, and WT+REF represent RNA from sunflower tumors containing only the wild type 780 gene, the 780 reference gene, or both, respectively. (**A**) S1 nuclease analysis of 5' deletion mutants. (**B**) S1 nuclease analysis of internal deletion mutants. Bands with sizes between 110 and 67 bp arise from readthrough transcription due to the divergence of the probe with test RNA at the position corresponding to the *Sal*1 linker in the test gene. The single band in lane ID-76/-12 is due to the 5' termini approximately 30 bp downstream of the minor promoter TATA box in its new position. Lane M contains pBR322 *Hpa*II-generated DNA marker.

bp
M
WT
REF
WT+REF
Δ-427
Δ-396
Δ-368
Δ-336
Δ-311
Δ-290
Δ-271
Δ-229
Δ-170
Δ-98
Δ-74
Δ+32
-PROBE-
-MIN-
WT+REF
ID-348/-290
ID-320/-290
ID-252/-171
ID-249/-98
ID-153/-171
ID-112/-98
ID-153/-98
ID-153/-37
ID-112/-37
ID-112/-12
ID-76/-74
ID-76/-37
ID-76/-12
ID-112/-74
122
110
90
76
67
MAJ
REF
A
B

Effect of 5'–Deletions on the Major Promoter Activity

Analysis of in vivo transcriptional activity of 5'–deletion mutants identified the 5'–boundary of the 780 gene promoter and indicated that a far upstream activator element was required for efficient expression. Results of the S1 nuclease mapping of the test and reference genes along with relative transcript levels (RTL) derived for the major and minor promoters are presented in Figs. 2–10A and 2–11. The activity plot of these results (Fig. 2–12) demonstrates the effect of systematic 5'–deletions on promoter function for both the major and minor start sites. This analysis indicated that the 5'–border of the 780 gene may lie between –476 and –427 bp (ca. 5% activity drop) with a distinct subcomponent of importance positioned near this border. The presence of this far upstream regulatory sequence was revealed by the dramatic reduction in RTL that occurred when the 31 bp region between –427 and –396 bp was deleted. Removal of these sequences reduced promoter activity by approximately 45%. It was also apparent that other components of the activator were located between position –396 and –229 bp. Deletion of this region of the activator resulted in a loss of activity from approximately 50% at –396 bp to 1% at –229 bp.

With the activator element deleted, the upstream element and TATA motif of the 780 gene was not enough to promote detectable levels of transcription. Removal of 5' sequences from –171 to +32 bp resulted in activities of 0.5% or less. This 200–fold drop in activity demonstrated the absolute requirement of the activator in obtaining wild type levels of transcription.

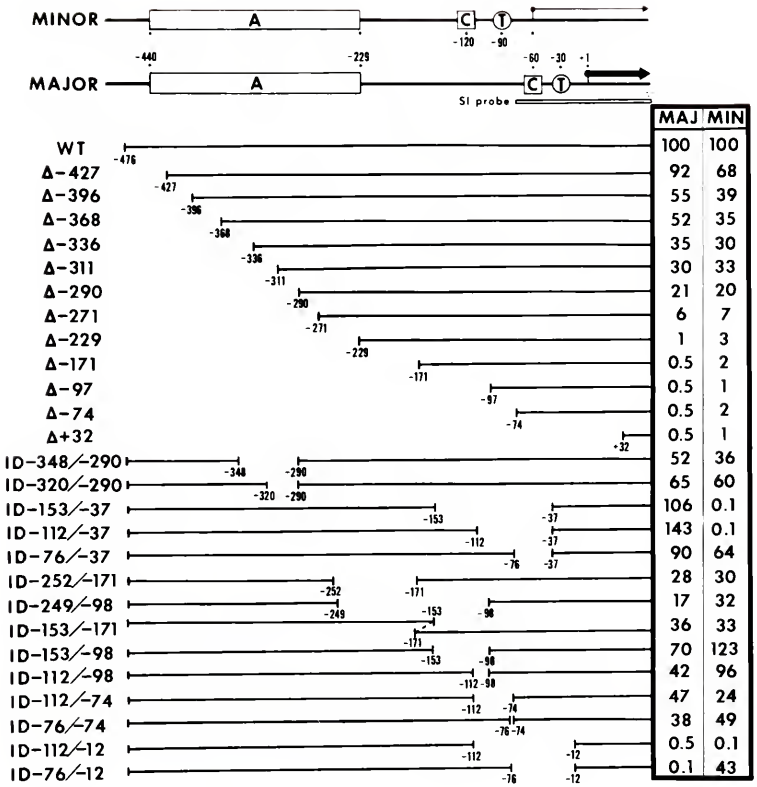


Figure 2–11. Diagrammatic representation of 780 gene mutations and RTLs. Deletion mutants shown here correspond to the deletions in Fig. 2–10. *Sal*1 linkers are represented by short vertical lines ending at each deletion site. The panel at the right shows the RTLs representing percentages of wild type activity for both the major and minor promoters. RTLs of >10 varied by 10% or less between experiments. Mutants that showed RTLs of <10 varied no more than ± 4% of wild type activity. Box A represents the activator element. Box C and circle T indicate the positions of the CAAT and TATA sequences of the major and minor promoters.

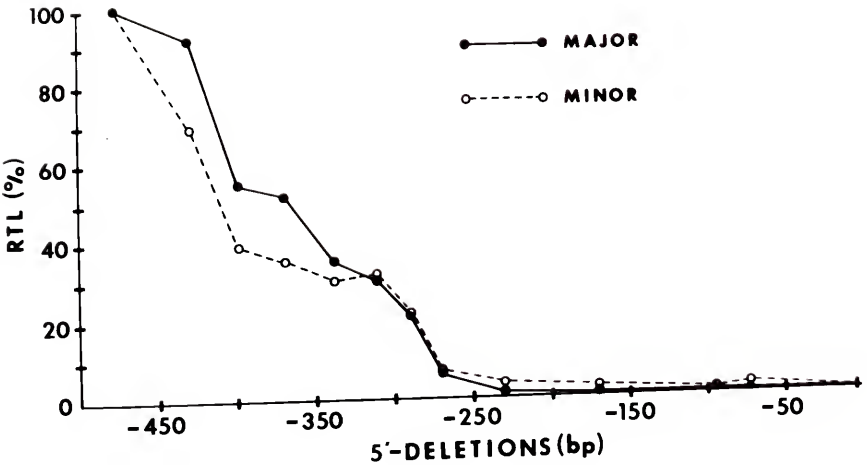


Figure 2–12. Effects of 5' deletions on transcriptional activity for both major and minor promoters. RTLs were plotted from the data shown in Fig. 2–11. Endpoints for the 5' deletions are numbered from the start site of the major transcript.

Effect of Internal Deletions on the Major Promoter

A total of 14 internal deletions and one 18 bp duplication mutant in the 5'–flanking region of the 780 gene were assayed for transcriptional activity (Figs. 2–10B and 2–11). These mutants were grouped into 4 classes according to the location of the mutation and their effect on the transcriptional activity of the major promoter. The first class of deletions (ID–348/–290 and ID–320/–290) were localized within the activator element delineated by the 5'–deletion series. Both deletions in this class reduced activity to between 52 and 65% of wild type levels which is similar in effect to the 5'–deletion at –396 bp. The results of these two internal deletions together with the 5'–deletion series suggested that the activator element is a single promoter element possibly having multiple subcomponents that function in concert.

A second class of internal deletions either stimulated or caused minor changes in transcriptional activity. Mutations in this class included ID–76/–37, ID–112/–37, ID–153/–37 and ID–200/–37 (this last deletion is presented in Fig. 3–4 and 3–5). A common feature of this group was the absence of sequences immediately upstream of the TATA box from –76 to –37. This particular deletion resulted in 90% of wild type activity but removal of more sequences upstream of –76 bp resulted in an increase of transcription from 106 to 140% of wild type. These deletions allow the activator element to come as close as possible to TATA removing parts, or all, of the upstream element sequences. The effects of completely removing this region (ID–200/–37) resulted in an increase in transcription relative to wild type indicating that sequences in the analogous position of the upstream element of animal genes were not absolutely required for normal levels of transcription.

Mutations located between position –229 bp (the putative 3'–boundary of the activator) and approximately –74 bp have been grouped in a third class.

Generally these mutations decreased transcriptional activity to below 50% of the wild type activity with the exception of ID–153/–98 which resulted in 70% activity. Deletions ID–252/–171 and ID–249/–98 were especially drastic in their effect, reducing activity to 17 and 28%, respectively. The severity of these two mutations may have been due in part to removal of sequences within the 3' end of the activator as well as part of the upstream element. Two small internal deletions ID–76/–74 and ID–112/–98 caused reductions in transcriptional activity to 38 and 42%, respectively, even though complete removal of this region (ID–112/–37) resulted in an RTL of 143%.

Internal deletions that remove the TATA comprise the last class of mutations and result in severe loss of promoter function. Deletions ID–76/–12 and ID–112/–12 showed less than 0.5% and 0.1% activity, respectively, suggesting that TATA is essential for 780 promoter function. These levels constrast strongly to the near wild type activities of the mutants ID–76/–37 and ID–112/–37 which still contained the TATA motif. Activity levels resulting from the removal of TATA were comparable to those obtained by removing the entire activator element with the 5'–deletion to position –229 bp suggesting that this motif was essential for normal activity.

Deletion Effects on the Minor Promoter

Analysis of the activity of the minor promoter suggested that it shares the activator element with the major promoter along with other structural similarities. The 5'–deletion activity plot in Fig. 2–12 showed that both the major and minor promoters have nearly identical profiles. This finding suggested that the activator element for the major promoter must also act in a similar fashion on the minor promoter. Mutants ID–112/–98 and ID–153/–98 are similar to those in the second class of the major promoter internal deletions in that

complete removal of sequences immediately upstream of the TATA (minor) resulted in either increasing by 23% or causing little change on the level of the minor transcripts. Mutant ID–153/–171 is similar to the third class of the major promoter internal deletions since a similar alteration (18 bp duplication) of sequences upstream of TATA also caused a decrease in activity of the minor promoter. Removal of the TATA (minor) also drastically decreased the minor transcripts to barely detectable amounts as shown by ID–112/–37 and ID–153/–37. However, reduction of the activity level of the minor promoter was not as drastic (reducing to 24%) with an internal deletion from –112 to –74 bp suggesting that some sequences immediately downstream of –74 bp are required for low level activity for this promoter.

The conservation in spacing between the TATA and the start of transcription was demonstrated for the minor start site. In Fig. 2–13 (lane WT) the start site for the minor transcript was mapped between –119 to –114 bp from the *Hpa*II site or approximately 60 bp upstream from the major start site. The insertion of an 8 bp *Sal*I linker between –17 and –15 bp relative to the minor start of transcription (ID–76/–74) resulted in the shifting of this start site 3 to 5 bp upstream. This maintained a distance of 27 to 29 bp between the third nucleotide of TATA (minor) and the start site of transcription. In a similar mutation (ID–76/–37), 39 bp including the minor cap site were removed and an 8 bp *Sal*I linker inserted. In this case the transcription start site did not appear as discrete as the wild type start and was located 20 to 40 bp downstream of TATA's (minor) new position. The level of activity for both of these mutations also decreased to 49% for ID–76/–74 and 64% for ID–76/–37. These results suggest that for the 780 gene minor promoter, the start of transcription is not primarily determined by sequences around the cap site, but most probably by

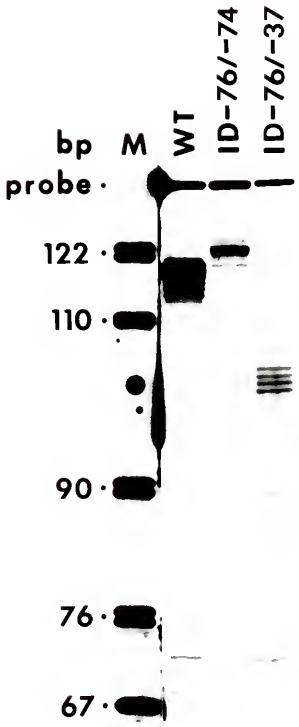


Figure 2–13. S1 nuclease mapping of the minor transcript. Approximately 15 µg of poly (A) RNA was hybridized to the S1 probe (Fig. 2–11). ID–76/–74 contains a 6 bp insertion between the TATA (minor) and the minor transcript cap site. ID–76/–37 contains a 31 bp deletion removing the minor cap site, thereby moving the TATA (minor) to a new position resulting in protected bands of about 80–90 bp in size. The 100 bp size bands represent a divergence of test gene transcripts (readthrough) at the position corresponding to the *Sal*I linker in the test gene. Lane M consists of a pBR322 *Hpa*II–generated marker.

TATA as in animal genes (83), and that the sequences from –76 to –37 bp are of some importance to the minor promoter transcription.

Conclusion

The transcriptional activity of a series of 5' and internal deletion mutants of the 780 gene from pTi15955 was assayed in order to determine the internal structure of the promoter. Precise quantitation of transcription was achieved by using a homologous reference gene acting as an internal standard while present in the same vector as the mutant test gene. Since the reference gene transcripts were shown to be distinguishable from the test gene transcripts, an accurate determination of transcript levels was possible for each mutation. Transcription factor saturation was avoided by using the T–DNA vector system in which a relatively few copies are integrated in the plant genome (106, 152, 175). The factor saturation problem usually accompanies viral–based vectors and DNA transformation systems which result in high template copy numbers (92) that can saturate out transcription factors. Variations in promoter activity due to the integration of T–DNA at random sites (43, 65, 73) was also avoided. Pooling 200–300 tumors for each mutation assay averaged the possible chromosomal location effects on the integrated T–DNA. With less than 10% variability of promoter activity between experiments, the sensitivity of this transcription expression system allowed the detection of discrete functional domains in the 5'–flanking region of the 780 gene.

Three functional domains were postulated based on the results of the 5' and internal deletion studies. These domains have been designated as the

activator, the upstream region and the TATA and appear to be similar to analogous elements in animal and viral promoters. Due to the limited number of mutations evaluated, the 5' and 3' border of the activator element and the boundaries of the upstream element are only approximate. The minor promoter was also shown to have three functional domains similar to the major promoter even though its own activity is much lower relative to the activity of the major promoter.

The 780 gene activator element is the most distal domain with respect to the start of transcription and is defined primarily by the effect of systematic 5'–deletions on the level of transcription (Fig. 2–12). The activator element was required for full transcriptional activity for both the major and minor promoters since its removal reduces the activity for both promoters 100 (major) to 30–fold (minor) less than wild type. The activator was also able to function closer to TATA than that present in the wild type gene. For example, the second class of internal deletions move the activator from 31 to 163 bp closer to TATA and showed either very little reduction in activity, or as much as a 40% increase over wild type transcription levels (Fig. 3–4).

A least characterized domain of the 780 promoter lies between the TATA (–37 bp) and the postulated 3' border of the activator (–229 bp). In numerous animal genes this region includes the domain designated as the upstream element which is generally positioned from –110 to –40 bp (35, 112). In general, this region is sensitive to slight disruptions in sequence composition and spacing (30, 102, 112). The region between TATA and the activator element of the 780 promoter seems to be analogous to the upstream element of animal promoters since small scale disruptions drastically reduce transcriptional activity demonstrating the limited spatial flexability of this element (30, 102). Larger deletions in the region, which included the CAAT box sequences

immediately upstream of TATA, were shown to stimulate transcription to as much as 40% over wild type levels, implying that these sequences may also impart a negative influence on the rate of transcription. With this in mind, small disruptions of this sequence should therefore interfere with the negative element's influence and allow the transcriptional rate to increase. Such mutations (ID–76/–74 and ID–112/–98), however, resulted in sizeable decreases in activity supporting the idea that the 780 upstream element provides a positive rather than a negative influence.

The second class of mutations suggested that the activator was able to substitute for the upstream element as long as the deletions positioned the activator close to TATA and removed large portions of the intervening sequences. Morelli et al. (109) observed a similar occurrence with the removal of sequences, which included the CAAT box, in the *rbc*S gene from pea that resulted in a nearly 2–fold increase in activity. The authors stated that this region may act as a negative element since its removal increased activity. An analogous event was demonstrated by the deletion of a similar region in the promoter of the sea urchin H2A histone gene (54) also resulting in a nearly 2–fold increase in transcription. Deletions of the promoter regions which include the CAAT box may not necessarily remove elements which have negative influences on gene expression. Alternatively, these two studies may suggest that other 5' elements far upstream, analogous to the 780 activator element, can be repositioned closer to TATA and increase the transcriptional activity possibly by optimizing interactions of factors involved with transcription. Therefore, the *rbc*S, histone and 780 genes may require some element to be located in the TATA–proximal position for efficient activity similar to the MT–1 gene previously mentioned (135).

In most eucaryotic promoters, a TATA motif is often positioned from 21 to 35 bp upstream of the start of transcription. In animal genes TATA is an essential component for activity of the promoter in some cases (30), but is only required for precise positioning of the transcriptional start site in others (6, 83). A clear demonstration of its role in positioning the start transcription of plant genes has not been reported. In this study the TATA element was absolutely required for the transcriptional activity of the major promoter. Its role in the positioning of the start of transcription was shown by the shift in the transcription start site obtained by insertion of sequences between the TATA (minor) and minor cap site, or by deletion of sequences between –76 and –37 bp (Fig. 2–13). These results imply that the function of the TATA in animal and plant promoters has been conserved throughout evolution.

In addition to TATA and CAAT motifs, no other sequences appear to be present in the promoter of the 780 gene (Fig. 2–14) which show strong homology with consensus sequences (112) commonly found in the promoters of eucaryotic genes. There are, however, four direct repeats (one 11 bp and three 7 bp in size) scattered throughout the 5'–flanking region. Three of these four repeats (*a*, *b*, *c*) are clustered near the 5'–border of the activator element. Deletion of these sequences, from position –427 to –396 bp, resulted in a sharp decrease in activity of both the major and minor promoters (Fig. 2–12) suggesting that some of these repeats may be critical to the activator function. Repeat *c* (5'–TTGAAAA–3') is located at three positions in the 5'–flanking region, whereas repeats *a*, *b* and *d* are present twice. Repeat *c* is also present once or twice in the 5'–flanking region in seven of the thirteen known genes of the octopine–type T–DNA (5). This same repeat is similar to the the sequence 5'–TTTCAAGGA–3' found in the 5'–flanking region of nopaline–type T–DNA genes (78). A seven out of nine base pair homology to this latter sequence is present

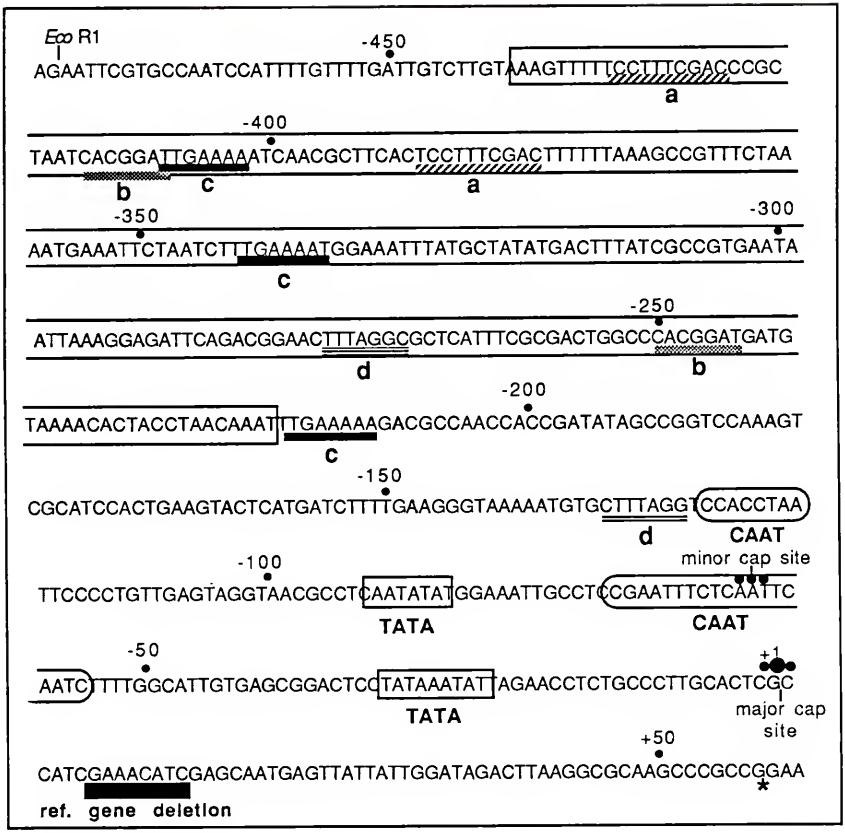


Figure 2–14. Sequence of the 5'–flanking region of the 780 gene. Nucletides are numbered from the major start of transcription. Solid circles above the sequence denote the 5' termini of the major and minor promoter. The large open box encloses the region of the activator element determined by the deletion mutation analysis. The short boxes refer to TATA sequences where designated. The CAAT motifs are circled for both the major and minor promoters. The asterisk denotes the site of the 5' ^{32}P end–label at position +60 bp of the hybridization probe which extends from position +60 to –74 bp. Individual direct repeats are designated **a**, **b**, **c**, and **d**.

twice in the 5'–flanking region of the 780 gene promoter centered at positions –294 and –148 bp. Although the significance of short repeated sequences in T–DNA genes is not known, short repeats have been shown to be involved with enhancer activity and to constitute sites of protein–DNA interaction in enhancers and upstream elements of animal genes (20, 71).

In conclusion, the functional domains within the promoter of the 780 gene have been partially characterized based on this limited deletion study. Of the three analyzed regions of the 780 promoter, the activator element seems to be the most enhancer–like since this element can be moved much closer than its wild type position relative to the cap site without disrupting the activity of the gene. Because the 780 gene is noninducible, the activator element may demonstrate some of the characteristics of a constitutive, or nonspecialized, enhancer element in plant promoters. A direct test of bidirectional function of the 780 activator in both the 5' and 3'–flanking regions of the gene is presented in the next chapter.

CHAPTER 3

ENHANCER PROPERTIES OF T-DNA PROMOTERS

Introduction

The existence of enhancers in T-DNA has not been reported in the literature although the results of the mutation studies summarized in Table 1 demonstrate the requirement for transcription of sequences 5' to the upstream element. As an example, sequences upstream to position -168 bp in the OCS gene are essential since no octopine is detected in tobacco tumors containing this gene with -168 bp deletion (82). Normal opine activity is restored when the deletion is only to -292 bp suggesting the requirement of sequences between -292 bp and -168 bp for OCS gene expression in tobacco. A second T-DNA gene that may be associated with an enhancer-like element is gene 4 which encodes an enzyme in the cytokinin pathway, dimethylallyl pyrophosphate transferase. Hooykaas et al. reported (Fallen Leaf Lake Conference on "The Genus *Agrobacterium* and Crown Gall," Lake Tahoe, Nevada, [September, 1986]) that a region, determined by internal deletions, between ca. -180 to -150 bp in the promoter of gene 4 was essential for activity. Previously, Lichtenstein et al. (91) demonstrated that a Tn5 insertion mutation of gene 4 at the -121 bp site still resulted in normal expression as assessed by the presence of normal tumor formation. Taken together, these results using gene 4 mutations suggest that sequences upstream of the -121 bp site could be repositioned nearly 5.5 kbp further upstream from the core promoter

(upstream element and TATA box) and the gene could still function normally. Thus there is evidence of required sequence 5' to the upstream element in tobacco tissue for both the OCS gene and gene 4. The location of promoter elements this far upstream, and the possibility of considerable flexibility in their spacing requirements suggests that these distal elements may have enhancer–like properties.

Based on the internal deletions in Chapter 2, the activator element of the 780 gene promoter was shown to exhibit some flexibity in spacing. The possibility that this activator element may have enhancer–like qualities was examined by functionally testing this element in both polarities upstream and downstream of a 5' deleted 780 gene. The deletion clone (pΔ–37 clone) contained only the TATA sequence and could not initiate any detectable transcription. Therefore, any change in transcriptional activity must be directly due to the presence of the activator element. The ability of the 780 gene activator element to function at locations further upstream of its normal wild type position was also examined. A *Hae*III fragment of approximately 600 bp in size from the replicative form of bacteriophage ΦX174 was used to separate the activator element from the TATA box present in the –37 deletion clone.

The existence of enhancer–like properties in the upstream regions of the OCS and agropine synthase genes (AGS) was also examined. The regions of the OCS and AGS promoters used in this study were located between the CAAT boxes and the T–DNA border sequences. These sequences were also placed 5' and 3' to the pΔ–37 5' deletion 780 gene and then introduced into sunflower using the double gene shuttle vector system. The utilization of the double gene vector system for these constructions provided accurate comparisons of the ability of these three T–DNA elements to activate transcription from a severly deleted 780 gene promoter.

Materials and Methods

Construction of 5' and 3' Bidirectional Orientations of the 780 Activator Region Relative to the 780 Gene pΔ–37 Deletion Clone

The 780 activator region was introduced into 5' and 3' locations of the pΔ–37 deletion clone of the 780 test gene. The activator element was inserted into either the *Bam*HI or *Hin*dIII sites of the double gene shuttle vector containing the 780 deletion mutant (Fig. 3–1). Such constructions allowed the activator element to be positioned in both polarities directly upstream from the TATA box, or approximately 200 bp downstream of the poly (A) addition site of the 780 gene. To acccomplish these constructions, the *Sal*I site of the –112 and the –200 bp 780 gene 3' deletion clones (see Fig. 2–4) were converted to *Bam*HI sites by linker addition (95). After digestion with *Bam*HI, the resultant fragments were then ligated into the *Bam*HI site of pW9–TD: Δ–37 and screened to determine polarities by digesting the final plasmid with *Eco*R1 and *Hin*dIII.

The activator element was also moved 538 bp further upstream than its position in the wild type gene by inserting the *Sal*I–linkered 603 bp *Hae*III fragment from ΦX174 (1173 to 1779 bp [129]) into the *Sal*I site situated between the activator element and the –37 bp deletion of the 780 gene. The 603 bp fragment was tested for its ability to affect transcriptional activity by being placed alone in both polarities directly upstream of the 780 gene Δ–37 clone. The polarities of the activator element upstream of the ΦX174 inserted fragment were determined by a triple digestion of the resultant plasmid with

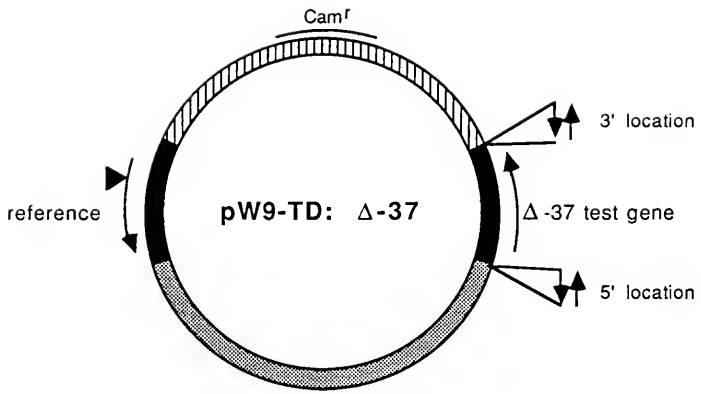


Figure 3–1. Introduction of enhancer fragments into the double gene shuttle vector. Enhancer fragments linkered with either *Bam*HI or *Hin*dIII were inserted in the 5' or 3' locations, respectively, of the 780 Δ–37 deletion clone in both polarities as described in the Materials and Methods. The hatched area is the *Sph*I–*Bam*HI fragment of pACYC184 (16). The solid black areas are the 780 reference gene and Δ–37 deletion 780 gene. The stippled area is the *Sph*I–*Bam*HI fragment from p233G (55). The black triangle denotes the *Taq*I deletion in the reference gene leader and cam^r denotes the chloramphenicol resistance gene.

*Eco*R1, *Pst*I and *Acc*I. The orientation of the ΦX174 *Hae*III fragment alone was determined by digesting the pW9/Δ–37:ΦX174 construction with *Acc*I.

To introduce the 780 activator element into the downstream location of the Δ–37 deletion clone, the *Eco*R1–*Sal*I fragment (-476 to –112 bp) from a –112 bp 3' deletion clone was modified by the addition of *Hin*dIII linkers (95). Following linker addition this fragment was then inserted into the *Hin*dIII site of the pW9–TD:Δ–37 in both orientations. The polarity of this element was determined by digestion of the final plasmid with *Eco*R1.

Construction the OCS and AGS Upstream Sequences Inserted in Both Polarities 5' and 3' to the Δ–37 deletion clone of 780

Sequences between the CAAT boxes and the respective 3' T–DNA borders from the OCS and AGS genes have been placed in both polarities, 5' and 3' to the Δ–37 780 gene deletion clone (Fig. 3–1). The *Bam*HI–*Acc*I fragment of the OCS gene (13,775 to 13,991 bp [5]) and the *Bam*HI–*Sal*I fragment of a 3' deletion mutant of the AGS gene (23,576 to 23,758 bp [5]) were modified by either *Bam*HI or *Hin*dIII linker addition (95). The OCS and AGS linkered fragments were then ligated into the *Bam*HI or *Hin*dIII site of pW9–TD:Δ–37 and screened by gel analysis in order to obtain both polarities. The polarity of the OCS fragment was determined by digestion of pW9–TD:Δ–37/OCS with *Hin*cII. The polarity of the AGS fragment was determined by first isolating either a *Pst*I–*Sal*I fragment, or an *Xba*I fragment for the 5' or 3' positions, respectively, from pW9–TD:Δ–37/AGS, then digesting the isolated fragments with *Rsa*I.

The double gene shuttle vectors with the 780 activator, OCS or AGS sequences in the 5' and 3' positions relative to the Δ–37 deletion clone were mobilized into *A. tumefaciens* Ag5260 and used to inoculate sunflowers seedlings as previously described. The RNA from 14–16 day tumors was extracted and analyzed by S1 nuclease hybrid protection mapping. RTL's were

determined from the cpm values of the radioactive bands that were cut out of the polyacrylamide gel and subjected to Cerenkov counting. The RTL values were calculated by the ratio of cpm values of the enhancer–element/test gene and the 780 reference gene divided by the ratio of cpm value of the wild–type 780 gene to its 780 reference gene.

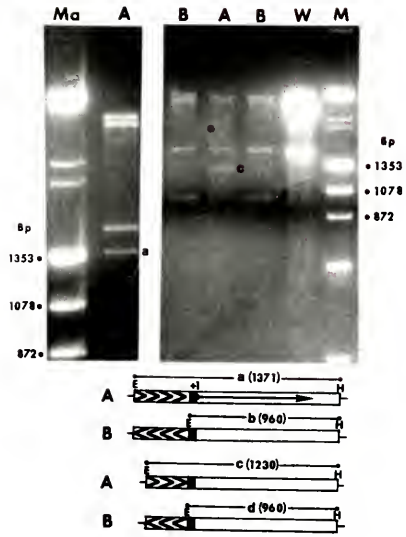
Results

Determination of Polarity of the Activator, OCS, AGS, and ΦX174 Fragments Relative to the Δ–37 Clone

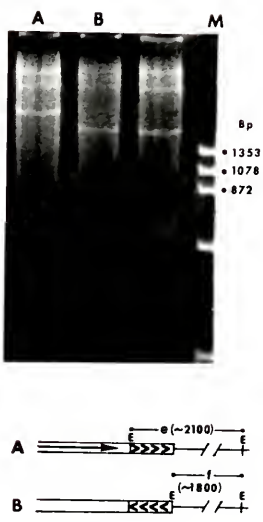
The polarity of the activator element inserted in both the 5' and 3' positions of a 5' deletion of the 780 gene was determined by restriction endonuclease digestions. Fig. 3–2 summarizes the results of the digestions involving the double gene shuttle vectors, pW9–TD: Δ–37, with the –476/–112, –476/-200, –476/-112: ΦX174 fragments, and the ΦX174 fragment alone. The polarities were designated A for the normal wild type polarity and B for the opposite polarity. After digestion with *Eco*R1 and *Hin*dIII, the predicted sizes for fragments "a" and "c" from the plasmid pW9–TD:Δ–37 containing the –476/–112 bp or –476/-200 bp sequences were 1357 and 1262 bp for the A orientation, respectively. The B orientation, by prediction, should result in 963 bp size fragments (bands "b" and "d") for both elements. The fragment –476/-112 in the 3' position should result in two bands, "e" (~ 2.0 kbp) for the A polarity and "f" (~ 1.8 kbp) for the B polarity, based sequence prediction. The actual sizes of bands " a–f" were similar to the predicted values (number 1, Fig. 3–2). Bands "l" and "k" representing both orientations of the ΦX174 fragment alone were also determined from sequence data (129) to be approximately 3.0 and 2.4 kbp which corresponded to sizes determined from the gel (number 3, Fig. 3–2). A triple digestion was needed to discern the polarity of the –476/-112

Figure 3-2. Determinaton of polarities of the 780 activator element and ΦX174 DNA fragments. Minipreparations of double gene shuttle vectors (with the 780 Δ-37 gene) plasmid DNA containing the -476/-112 bp or -476/-200 bp fragments of the activator element or the ΦX174 603 bp *Hae*III fragment in either orientation (lanes A and B for each fragment) were digested with restriction enzymes and separated by gel electrophoresis. (**1**.) The -476/-112 bp and -476/-200 bp fragment in the 5' position (780 5') and the –476/–112 bp fragment in the 3' position (780 3'). (**2**.) The -476/-112 bp fragment upstream of the ΦX174 fragment. (**3**.) The ΦX174 fragment in the 5' position. A and B are the normal and opposite polarities, respectively. All sizes are in basepairs determined directly from the gels. The >>> in the open box denotes the polarities of the 780 activator element fragments while the hatched box refers to the ΦX174 fragment. The small black box represents the region between -37 and +1 bp of the 780 promoter. The arrow is the 780 transcript. Lane Ma contains λ *Hin*dIII-generated markers and ΦX174 *Hae*III-generated markers while lane M only contains the ΦX174 *Hae*III-generated markers. Legend: Ac, *Acc*I; E, *Eco*R1; H, *Hin*dIII; P, *Pst*I.

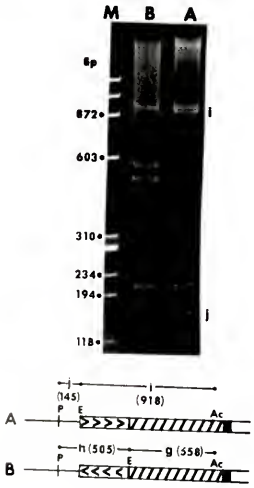
1. 780 5′



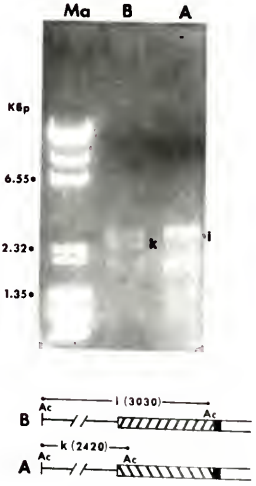
780 3′



2. 780/ΦX174



3. ΦX174



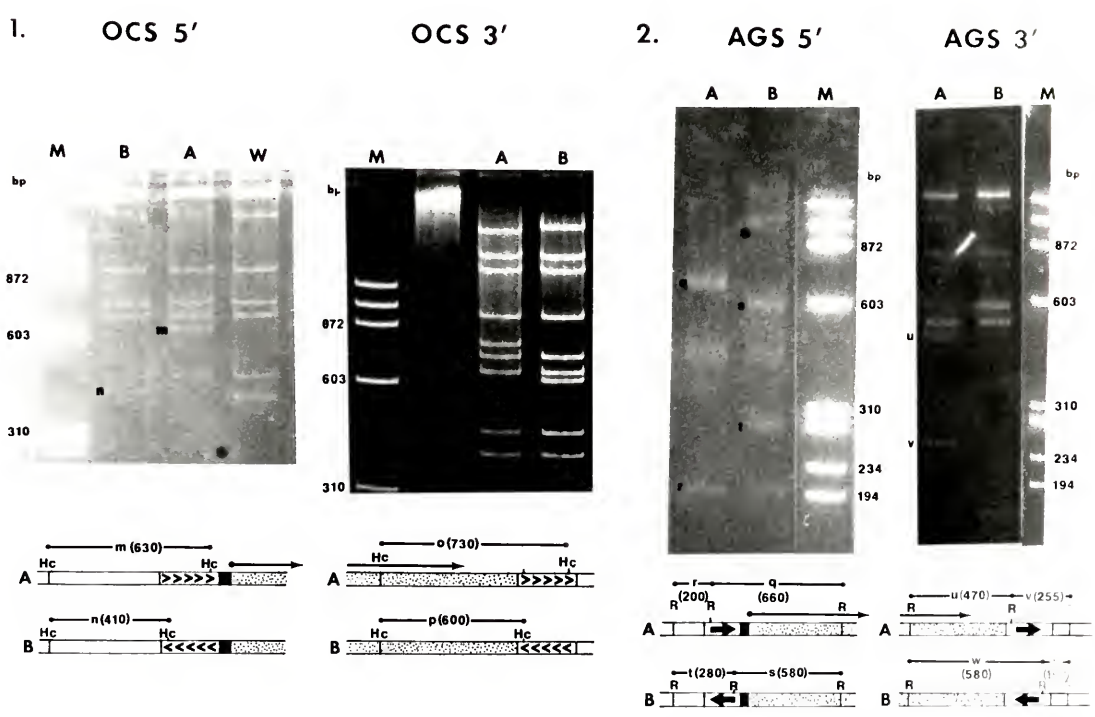
fragment relative to the ΦX174 fragment and the Δ–37 clone. For this construction the ΦX174 fragment was in the A orientation (determined by bands "g" and "i"). The sizes of the –476/-112 fragment (bands "h" and "j") was predicted to be 513 and 149 bp for both polarities which were similar to the band sizes based on the electrophoretic gel (number 2, Fig. 3–2).

The polarities of the OCS and the AGS fragments relative to the Δ–37 clone in both the 5' and 3' positions were likewise determined by comparing the size predicted with the actual sizes determined by the gel electrophoresis (Fig. 3–3). After *Hin*cII digestion, the OCS fragment in the A orientation in the 5' and 3' locations should yield fragments with sizes of 610 and 732 bp, respectively. The B orientation should yield fragments of 402 and 577 bp for the 5' and 3' positions, respectively, based on sequence prediction. Portions of the double gene shuttle vector plasmid containing the Δ–37 clone and the AGS fragment were first isolated then digested with *Rsa*1 as described in Materials and Methods. The polarity of the AGS fragment was determined by sequence prediction of the bands "r–s" (band "r" was a doublet), produce fragments having sizes of 190 and 652 bp for the 5' A orientation and 294 and 548 bp for the 5' B orientation. The 3' location resulted in the bands "u–x" with predicted sizes of 441 and 262 bp for the A orientation and 550 and 153 bp for the B orientation.

Effects of the Activator Region in Different Polarities on Δ–37 Clone of the 780 Gene

The function of the 780 activator element in promoting transcription of the Δ–37 clone while in the upstream position was assessed by S1 nuclease mapping. An autoradiogram of the S1 nuclease mapping gel shows the results of the transcriptional activity of the deletion clone with the activator element present in both polarities (Fig. 3–4A). A summary of the constructions

Figure 3-3. Determination of polarities of the OCS and AGS DNA fragments. Minipreparations of double gene shuttle vectors (with the 780 Δ-37 gene) plasmid DNA containing the OCS and AGS fragments in either orientation were digested with restriction enzymes and separated by gel electrophoresis. (**1**.) The OCS fragment in either polarity (lanes A and B) in the 5' and 3' positions. (**2**.) The AGS fragment in either polarity (lanes A and B) in the 5' and 3' positions. The >>> in the box denote the polarities of the OCS fragment while the short thick arrows denote the polarities of the AGS fragment. The stippled area is the 780 gene with the long thin arrow representing the 780 transcript. The black box denotes the region of the 780 promoter between -37 and +1 bp. All sizes are in basepairs and determined directly from the gels. Lane M contains ΦX174 *Hae*III-generated markers while lane W contains the double gene shuttle vector (with the 780 Δ-37 gene) digested with *Hin*cII. Legend: Hc, *Hin*cII; R, *Rsa*I.


1.
OCS 5′
OCS 3′
M
B
A
W
bp
872
603
310
m (630)
n (410)
o (730)
p (600)
Hc
2.
AGS 5′
AGS 3′
r
q
(200)
(660)
t (280)
s (580)
u (470)
v (255)
w
(580)
R

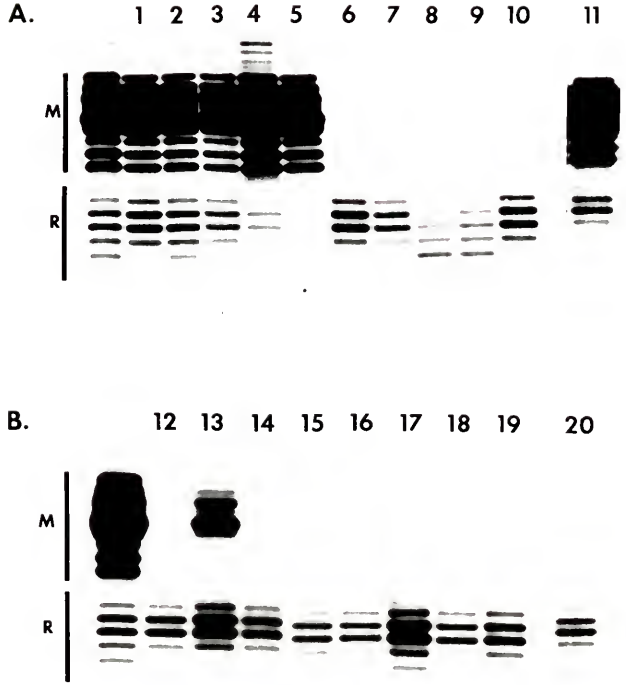


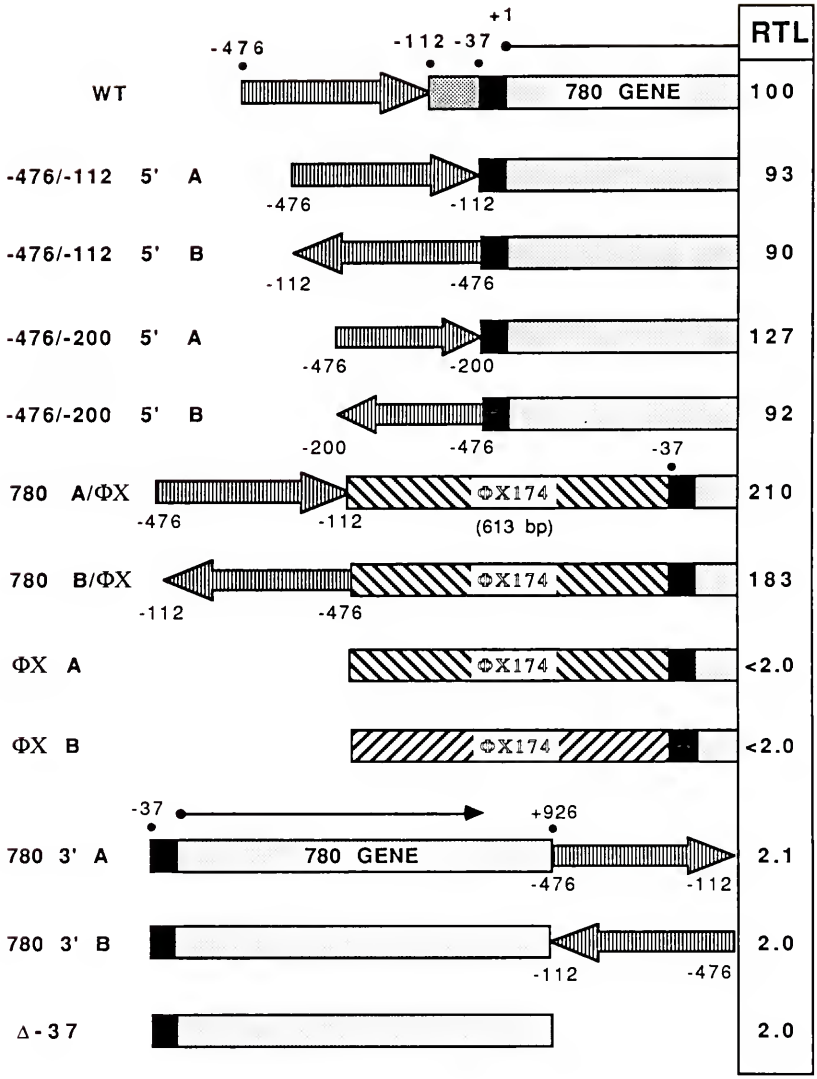
Figure 3–4. Autoradiograph of S1 nuclease analysis of the 780 Δ–37 gene containing the various T–DNA fragments in different locations. Approximately 15 µg of poly (A) RNA was used for all lanes. The unmarked lanes contain RNA from sunflower tumors with the reference and wild type genes. (A.) Lane 1, –476/-112 5' A; lane 2, –476/–112 5' B; lane 3, –476/-200 5' B; lane 4, 780 (-476/-112) A/ΦX; lane 5, 780 B/ΦX; lane 6, ΦX A; lane 7, ΦX B; lane 8, 780 3' A; lane 9, 780 3' B; lane 10, 780 Δ–37 gene; and lane 11, –476/-200 5' A. (B.) Lane 12, OCS 5' A; lane 13, OCS 5' B; lane 14, AGS 5' A; lane 15, AGS 5' B; lane 16, OCS 3' A; lane 17, OCS 3' B; lane 18, AGS 3' A; lane 19, AGS 3' B; and lane 20, 780 Δ–37 gene. "M" refers to the 780 major transcripts and "R" designates the reference gene transcripts. The hybridiztion probe and conditions are the same as in Fig. 2–10. Lanes 11 and 20 are from another autoradiogram.

involving the 780 activator element and the relative transcript levels is shown in Fig. 3–5. Using the construction –476/-112 5' A, in which the 780 promoter between positions –112 bp and –37 bp was replaced by 25 bp of a portion of the polylinker from pUC–19, activity was reduced slightly to 93% of wild type levels. When the –476/-112 bp fragment was placed upstream in the opposite polarity, the activity was reduced further to a RTL value of 90%. Removal of 163 bp of internal promoter sequences of the 780 gene between –200 to –37 bp (-476/-200 5' A, Fig. 3–5) and replacing it with the 25 bp pUC–19 polylinker DNA resulted in 127% of transcriptional activity. However, reversing the orientation of the –476/-200 bp fragment (-476/-200 5' B) reduced transcription by 35%. In summary, the activator region in both polarities efficiently promoted transcriptional activity of the deleted 780 gene, which contained only TATA box sequences, to approximately wild type levels.

The activator domain between –476 and –112 bp of the 780 promoter was also able to activate the Δ–37 deletion clone when separated from the –37 bp site by 603 bp of ΦX174 sequences. Moving the –476/-112 fragment to this new position resulted in a 2–fold increase in transcription relative to the wild type level. Reversing the polarity of the activator element in the same location slightly reduced this elevated activity to 183% of normal levels. The ΦX174 fragment alone in either orientation was unable to stimulate significant transcription resulting in activity levels similar to the Δ–37 deletion clone alone. Evidently the activator element does not have stringent spatial requirements for functioning in the 5' position since this element could promote high levels of transcription from –650 bp to –37 bp upstream of TATA.

The activity level of the construction, –476/-112 5' A, differs significantly from the activity of a similar construction, ID–112/-37 (Fig. 2–11) containing an internal deletion of sequences between –112 to –37 bp

Figure 3-5. Schematic of 780 activator element in various positions relative to the 780 Δ–37 gene. The activator constructions correspond to the S1 nuclease analysis in Fig. 3-4A. The vertical-lined arrows represent either the –476/–112 bp or –476/–200 bp fragments from the 780 promoter as designated. The grey and black boxes denote the regions of the 780 promoter from –112 to –38 bp and –37 to +1 bp, respectively. The stippled region represents the 780 gene from +1 bp to approximately 200 bp downstream of the poly (A) addition site. The hatched box is the ΦX174 *Hae*III 603 bp fragment with 10 bp *Sal*I linkers added. The panel at the right shows the RTLs representing percentages of 780 wild type gene activity (WT). Usually the RTLs varied to within ± 10% of the percentage value whereas the RTLs of 2.1 or less varied only ± 0.9 percentage points between experiments.


RTL
+1
-476
-112 -37
WT
780 GENE
100
-476/-112 5' A
-476
-112
93
-476/-112 5' B
-112
-476
90
-476/-200 5' A
-476
-200
127
-476/-200 5' B
-200
-476
-37
92
780 A/ΦX
ΦX174
-476
-112
(613 bp)
210
780 B/ΦX
ΦX174
-112
-476
183
ΦX A
ΦX174
<2.0
ΦX B
ΦX174
<2.0
-37
+926
780 3' A
780 GENE
-476
-112
2.1
780 3' B
-112
-476
2.0
Δ-37
2.0

which was previously shown to be 140% of wild type levels. The ID–112/-37 mutation was constructed with a 6 bp *Sal*I linker joining the two deletion endpoints. Due to an additional 19 bp present between –112 bp and –37 bp deletion sites in the –476/-112 5' A construction, this latter mutant was 1/3 less active in transcription compared to the activity of the ID–112/-37 deletion mutant. It is unclear why the additional 19 bp causes a reduction in activity since the –476/-112 bp fragment was shown to function quite well when it was positioned 613 bp further upstream.

The 780 activator was not, however, able to stimulate any detectable level of transcription when placed downstream of the Δ–37 deletion clone. Positioning the –476/-112 bp fragment approximately 200 bp downstream of the poly (A) addition site resulted in no activity in either polarity. This result is similar to the studies of the light–inducible enhancers from *rbc*S and Cab genes which do not activate transcription of the NOS promoter while in the downstream location (145, 153). Nevertheless the 780 promoter fragment from –476 to –112 bp seemed to have some enhancer–like qualities, functioning upstream regardless of position and orientation even though it could not function downstream of the gene.

The start site of the 780 major promoter remained constant in all of the constructions in which the test gene transcripts could be detected by the S1 nuclease mapping analysis (Fig. 3–4). The transcript levels from the minor promoter were not examined since the DNA probe used for the S1 nuclease mapping analysis diverges with the test gene transcripts at the –37 bp site where *Sal*I linkers were added. However, some low level transcripts starting upstream of –37 bp in the activator constructions immediately uptream of Δ–37 bp clone seemed to occur but only represented approximately 1–2% of the

major promoter activity (data not shown). These upstream start sites were not investigated further.

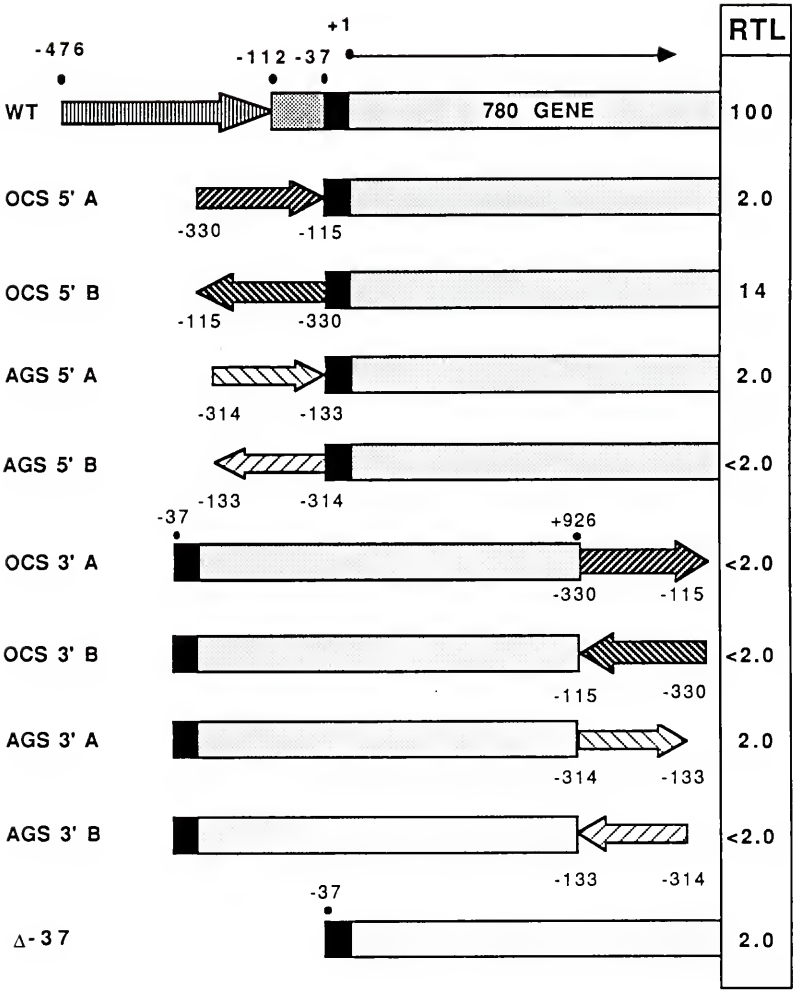
Evaluation of Enhancer-like Properties of Far Upstream Regions of the OCS and AGS Promoter

The promoter region immediately upstream of the CAAT boxes of the OCS and AGS genes were examined for their ability to promote activity of the Δ-37 bp deletion mutant of the 780 gene. The OCS promoter region from positions -330 to -115 bp, and the AGS promoter region from positions -314 to -155 bp (relative to their cap sites) were placed separately in both polarities upstream and downstream of the 780 gene Δ-37 clone. The autoradiographs of S1 nuclease mapping gels involving the OCS and AGS constructions are shown in Fig. 3-4B with the calculated RTL values for each construction shown in Fig. 3-6. Placing these heterologous promoter fragments either immediately upstream of the Δ-37 clone, or approximately 200 bp downstream of the poly (A) addition site, in both polarities produced levels of activity comparable to the Δ-37 deletion alone with one exception. The OCS fragment in the B orientation resulted in 14% of activity relative to the 780 wild type gene activity level. The activity in the B orientation was a 7-fold increase over the activity level (2%) of the OCS fragment in the normal A orientation.

Conclusion

The activator region of the 780 gene was demonstrated to possess some characteristics which are also shared by enhancer elements. It promotes transcription when placed in different 5' positions relative to the 780 gene. The activator region, in contrast to enhancer elements, was unable to function in the downstream location. The transcriptional activity was elevated

Figure 3-6. Schematic of the OCS and AGS fragment in various positions relative to the 780 Δ-37 gene. The OCS and AGS constructions correspond to the S1 nuclease analysis in Fig. 3-4B. The dark hatched arrows represent the the OCS fragment (-330 to -115 bp relative to the original OCS gene cap site) whereas the light hatched arrows represent AGS promoter fragment (-314 to -133 bp relative to the original AGS gene cap site). "WT" refers to the wild type 780 gene with its components as described in Fig. 3-5. The panel at the right shows the RTLs representing percentages of 780 wild type gene activity. The RTLs varied to within ± 0.8 percentage points between experiments.


RTL
+1
-476
-112 -37
WT
780 GENE
100
OCS 5' A
2.0
-330
-115
OCS 5' B
14
-115
-330
AGS 5' A
2.0
-314
-133
AGS 5' B
<2.0
-133
-314
-37
+926
OCS 3' A
<2.0
-330
-115
OCS 3' B
<2.0
-115
-330
AGS 3' A
2.0
-314
-133
AGS 3' B
<2.0
-133
-314
-37
Δ-37
2.0

approximately 2–fold over wild type levels when this element was placed further (613 bp) upstream by inserting a ΦX174 fragment between the activator element and the –37 bp site. This latter construction also demonstrated that the 780 gene does not require an upstream element for wild type levels of activity since the upstream element could be replaced by ΦX174 sequences and still retain high activity. By sequence inspection, the –476/-112 bp and –476/-200 bp fragments in both polarities contain CAAT box–like sequences that may replace the wild type CAAT boxes which were removed during the constructions of the promoter mutations. Whether these CAAT box–like sequences actually contribute to the transcriptional regulation is still unresolved.

Aside from the activator element acting in a constitutive manner, this element shares some properties of the *rbc*S light–inducible enhancer–like element. Both of these elements are able to promote activity in either polarity but apparently only in the 5' location. Also neither the 780 gene nor the *rbc*S gene seem to require sequences around the CAAT box since deletion of these TATA–proximal regions result in elevation of the transcriptional activities. It is possible that plant enhancer-like elements may only work in the 5' location of a gene since a demonstration of any such activities in the 3' position has not been made.

The OCS and AGS sequences upstream of their CAAT boxes, in general, did not efficiently activate transcription of the Δ–37 bp 780 deletion clone except for one construction involving the reverse polarity of the OCS fragment. This particular construction stimulated transcription 7–fold over background levels but was still only 1/7 of the 780 gene wild type level. An explanation for the B orientation of the OCS fragment promoting this increase in transcription when compared to the A orientation could be due to the presence of regions of

homology of the OCS fragment in the B orientation with the 780 promoter region. The 780 promoter does show such homologies including a 14 bp region (also includes the distal *a* repeat) from –435 to –421 bp of the 780 promoter (see Fig. 2–14) that is similar to a 14 bp sequence (with 2 bp mismatch) present in the reversed OCS fragment positioned from –220 to –234 bp relative to the cap site of the 780 Δ–37 deletion clone. Another 11 bp sequence present in the OCS fragment in the B orientation is positioned –130 to –141 bp relative to the cap site of the Δ–37 deletion clone which is homologous to a similar sequence in the 780 promoter positioned at –124 to –113 bp (allowing for 2 bp mismatch).

In addition, the 780 gene repeat *c* has nearly 80% homologies to a few regions of the OCS sequence in both polarities. Two sites of homology can be found in the OCS sequence in the B orientation at positions –267 and –180 bp relative to the deleted 780 gene cap site whereas one site can be found in the A orientation at position –194 bp relative to the Δ–37 cap site. Sequences of AGS and ΦX174 fragments or the T–DNA region upstream of all deletion constructions do not show any obvious homologies. The reason for the OCS fragment only working in one orientation is still unclear, since it contains the sequence homologies to the the activator element which was able to work in both polarities. Therefore further examination is still needed of whether these sequences that are homologous to the OCS fragment in the B orientation are important for transcriptional activity.

It is also possible the OCS and AGS genes may have enhancer–like elements present in their promoters but these elements are only capable of functioning in the presence of both an upstream element and TATA box. The only previous demonstration of an enhancer–like element within the OCS promoter was reported by A. J. Peacock et al., at the First International Symposium on Plant Molecular Genetics in Savannah, Georgia, (October

28–November 2, 1985). They reported that a fragment from the OCS promoter, located upstream of the CAAT box, was able to promote inducible activities while positioned in either polarity at the –145 bp site of *Adh–1* gene of maize. In this case, the –145 bp promoter of *Adh–1* may still include its own functional upstream element which may possibly fulfill the requirement for an upstream element by the putative OCS enhancer element for activation of transcription. The 780 activator region seems to differ from the OCS or AGS fragments in this respect by not having a strong requirement for an upstream element for activator function.

Another possibility of why the OCS or AGS fragment did not efficiently activate transcription of the 780 Δ–37 bp deletion clone was that these promoter sequences may only effectively activate their own core promoter (upstream element and TATA box) regions and not certain classes of heterologous core promoters. An example of such specificity of interaction is seen in the enhancers of the immunoglobulin genes. Both the κ chain and the heavy chain enhancers stimulated their own promoters 20–fold when compared to their effect on SV40 (lacking its enhancer region) and metallothionein (including nearly 2 kbp of promoter sequence) (48). The TATA box of the HSV tk gene also preferentially functions with its own upstream element (17). It is therefore possible that the OCS and AGS fragments may have a preference for particular promoter sequences in their activation of transcription.

In conclusion, the 780 enhancer–like element may be different to other similar elements in other T–DNA genes, like the OCS and AGS genes. It is found further upstream than elements required for activity in the other characterized T–DNA genes, and it can function in both polarities without the presence of the CAAT box. By analogy with animal enhancers, the 780

activator is likely to interact with specific transcription factors in the regulation of transcription. The possibility of such an interaction was examined using various plant nuclear extracts in the next chapter.

CHAPTER 4

IN VITRO NUCLEAR FACTOR BINDING TO THE 780 ACTIVATOR ELEMENT

Introduction

Enhancers in animal genes mediate their control of transcription through the specific binding of nuclear *trans*–acting factors. This interaction is able to engage other promoter elements near the initiation site over a considerable distance. The activator element of the 780 gene was previously shown in chapter 3 to have enhancer–like characteristics. This element residing nearly 200 bp from the TATA box and having a bidirectional activity, is likely to bind nuclear factors and initiate transcription in a fashion similar to the enhancers characterized in animal and viral genes.

Since the 780 gene was previously shown to be transcribed in tobacco as well as sunflower tissues, these two plants species may contain similar transcriptional factors interacting with the 780 promoter elements. Studies of the immunoglobulin (137, 169), α_1–globin (20), and the proto–oncogene c–*fos* genes (50) in animals utilized an electrophoretic mobility shift assay to demonstrate specific protein–DNA interaction. The electrophoretic mobility shift assay is useful in identifying regions of the promoter where specific binding occurs (20, 50, 169) as well as in the screening for the presence of specific DNA binding factors from different tissues (137). In this study the activator element was shown to bind specifically to factors from plant crude

nuclear extracts originating from two different plant species. This interaction was first determined by electrophoretic mobility shift assay and then examined more closely by in vitro DNase I digestion protection analysis. The electrophoretic mobility shift assay detects specific factor binding to a labeled DNA fragment which causes a retardation of mobility relative to the free DNA fragment when separated on a low percent acrylamide gel. The assay is a relatively simple, allowing one to characterize binding under a variety of conditions as well as roughly map areas of a promoter where specific interactions with nuclear factors occur.

The DNase I protection analysis permits a nucleotide level resolution, of regions protected from digestion by factor binding. This technique can indentify short sequences recognized by certain nuclear factors and has been utilized in the determination of the sequence of the Sp1 binding site (34). It may be possible to obtain a more complete picture of the 780 gene promoter function by correlating functional domains identified by deletion analysis with those regions involved in specific factor binding.

Materials and Methods

Preparation of Crude Nuclear Extract from Plant Plumules

Preparation of crude nuclear extracts from sunflower and soybean were performed from a procedure (169) as follows. Approximately 100 g of plumules from either one–week old sunflower (*H. annuus* cv. Large Grey) or etiolated soybean seedlings (*Glycine max*) were placed in 2X v/wt of ice cold solution I (10 mM Hepes [pH 7.9], 0.3 M sucrose, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT and 0.5 mM phenylmethylsulfonyl flouride (PMSF, Sigma). Most of the extraction procedure was performed in the cold room at 4°C. The plant mixture was homogenized for 1–2 minutes using a Tekmar

Tissuemizer at high speed, then filtered through mira cloth (Calbiochem) reinforced with cheescloth. The nuclei from the filtrate was collected by centrifugation at 2000 Xg for 10 minutes at 4°C. The supernatent was discarded and the pellet gently resuspended in 20 ml of solution I. The resuspended nuclei were mixed with 13 strokes using a Dounce homogenizer (2 strokes are a down and up movement) then collected again by centrifugation at 2000 Xg for 10 minutes at 4°C. The supernatant was discarded and the loose nuclear pellet was resuspended gently in 7 ml of solution II (10 mM Hepes [pH 7.9], 450 mM NaCl, 1.5 mM $MgCl_2$, 0.1 mM EGTA, 0.5 mM DTT, 0.5 mM PMSF and 5% glycerol) using a stirring rod every 5 minutes while on ice. This gentle stirring was carried out for 30 minutes to allow nuclear factors to diffuse out of intact nuclei. Nuclear material and any remaining debris were removed from this mixture by centrifugation at 100,000 Xg for 1 hour at 4°C. The supernate was dispensed into aliquots, frozen in liquid N_2, and stored at –70°C. Protein concentrations were determined by the method of Bradford (12).

Electrophoretic Mobility Shift Assay

Electrophoretic mobility shift assays using the activator region of the 780 gene were performed by a modified method of Singh et al. (147). Approximately 0.5–1.0 ng of an end–labeled fragment (-476 to –200 bp of the 780 gene promoter unless otherwise specificied) was incubated with 2.5 µl of soybean extract in a final volume of 26 µl containing 16 mM Hepes [pH7.9], 60 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, and 20% glycerol. A typical binding assay involving the sunflower nuclear extract was carried out at room temperature in a final reaction volume of 20 µl consisiting of 10 mM Tris–HCl (pH 7.5), 1 mM DTT, 1 mM EDTA, 5% glycerol and 2.5 µl of the sunflower extract present in solution II. The amount of poly (dI·dC) DNA, competitor DNA,

noncompetitor DNA, Triton–X–100 (Sigma), and $MgCl_2$ were varied as indicated (see Results).

After 30 minutes, the binding reaction was fractionated on a 4% polyacrylamide gel (30:1, acrylamide to bis–acrylamide) with recirculated buffer consisting of 6.7 mM Tris–HCl (pH 7.5), 3.3 mM Na acetate, and 1 mM EDTA. The gel was prerun for 30 minutes at 150 volts (11 volts/cm) at room temperature before loading the samples which were then run at the same voltage for 4 hours. After the run, the gel was fixed in 10% acetic acid, 10% methanol solution for 20 minutes, washed with distilled water for 15 minutes, then dried under vacuum on a 3MM Whatman filter using a heated slab gel drier. The dried gel was exposed to XRP–5 X–ray film (Kodak) in the presence of an intensifier screen at –70°C for 2–5 days.

DNase I Protection Assay

DNase I protection assays were performed on the activator region of the 780 gene using the nuclear extract from soybean plumules. Approximately 1–2 ng of 3' end–labeled DNA fragment (–476 to –200 bp of the 780 gene promoter) with 100 ng of ΦX174 *Hae*III digested DNA as carrier were incubated with 0.5–3 µl of solution II containing the crude nuclear extract. The DNA was added to a final volume of 44 µl of 0.7 mM Hepes (pH 7.9), 30 mM NaCl, 0.1 mM $MgCl_2$, 0.03 mM DTT, 0.03 mM PMSF and 3% glycerol, and incubated for 30 minutes at room temperature. After binding, 5 µl of DNase I buffer was added to final concentrations of 50 mM Na acetate (pH 6.5), 10 mM $MgCl_2$, and 2 mM $CaCl_2$ containing freshly diluted DNase I (Bethesda Research Laboratories) at a final concentration of 25 ng/ml and incubated for 3 minutes at room temperature. The optimum amount of DNase I was determined emperically for the best distribution of digested DNA bands. The DNase I reaction was terminated by

adding 1/10 volume of 200 mM EDTA and 10 µg of tRNA before extracting with phenol:chloroform: isoamyl alcohol (25:24:1). The DNA fragments were precipitated by the addition of 1/9 volume of 3M Na acetate (pH 5.2) and 2.2 X volume of 95% ethanol. The resuspended pellets were analyzed on an 8% polyacrylamide gel with 7 M urea, and exposed to XAR–5 X–ray film for 3–7 days at –70°C with an intensifier screen.

Results

Analysis of Nuclear Factors Complexed with 780 Activator Region

The electrophoretic mobility shift assay was used to assess the ability of various fragments of the 780 activator region to form specific DNA–protein complexes when incubated with crude nuclear extracts from sunflower or soybean seedlings. Specific conditions of the binding reaction were varied to obtain optimum interaction. Homologous and nonhomologous DNA was also used for competition in the in vitro binding reaction with the end–labeled probe to determine the specificity of the DNA–protein complex.

Four 3' end–labeled probes from different regions of the 780 activator region were incubated with sunflower extracts and examined by the electrophoretic mobility shift assay. Low intensity bands with reduced mobilities were detected while using 3 of the 4 fragments which included the –476/––347, –382/–248, and –310/–189 bp regions (Fig. 4–1). The –427/–290 bp fragment apparently did not bind to the factors since no distinct band was seen to shift. Factor interaction with the activator region seemed to be localized to two regions: one from –476 to –427 bp and the other from –290 to –189 bp. The apparent lack of binding of the –427/-290 bp fragment may reflect

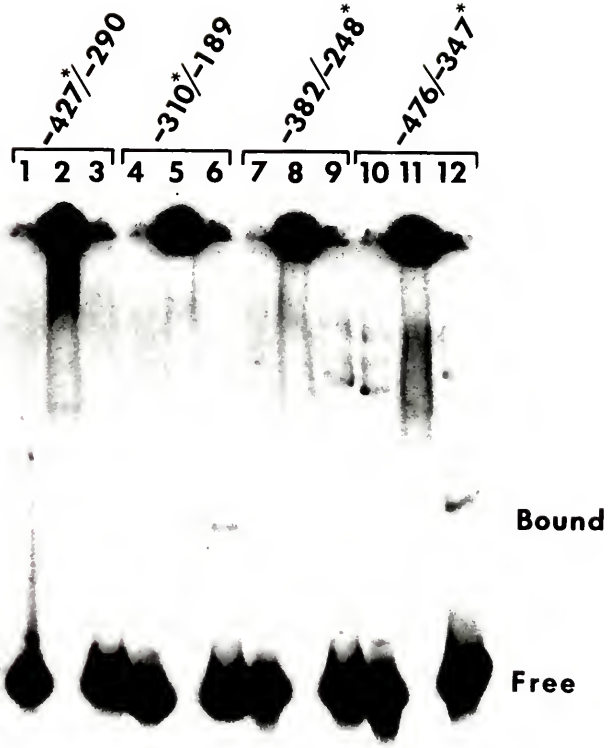


Figure 4–1. Electrophoretic mobility shift assay of various 780 activator element fragments using sunflower extract. These activator fragments were excised from either 5' or 3' 780 deletion clones and 3' end–labeled at the *Sal*I sites (*). Approximately 1 ng (3000 cpm) of these labeled fragments was incubated with 2.5 µl of crude sunflower extract as described in Materials and Methods. Lanes 1, 4, 7, and 10 contain free DNA fragment. Lanes 2, 5, 8, and 11 contain labeled fragment incubated with the extract. Lanes 3, 6, 9 and 12 contain labeled DNA incubated with extract and 5 µg of poly (dI·dC) DNA. Bound refers to the specific protein–DNA complex and free refers to the unbound labeled fragment.

a true absence of factor binding, or may simply be due to a lack of optimization of in vitro binding parameters.

Additional electrophoretic shift assays were conducted using the –476/-200 bp fragment in order to acheive an optimization of in vitro binding (Fig. 4–2). Multiple protein–DNA complexes were evident by the decrease in mobility of the end–labeled fragment in the presence of sunflower extract. A multiple banding pattern was also seen using cauliflower crude nuclear extract (prepared in a similar manner as the sunflower and soybean extracts) bound to the activator region (data not shown). Addition or removal of the binding salt solution, or the presence of Triton–X–100 in the binding reaction did not seem to affect the pattern of shifted complexes to any significant level (lanes 3–5, Fig. 4–2). The intensities of the bands, however, were highly dependent on the amount of extract and poly (dI·dC), or the presence of a divalent cation in the binding reaction. The addition of more extract and poly (dI·dC) DNA significantly decreased the complexes' mobility. Also, adding 10 mM $MgCl_2$ to the binding reaction caused four distinct bands to form with the three slowest bands designated as "a," "b," and "c" (lane 7, Fig. 4–2). Again, adding more extract and poly (dI·dC) DNA with $MgCl_2$ (lane 9) slightly decreased the mobility of bands "a," "b," and "c" while increasing their level of intensity. The $MgCl_2$ was therefore added to subsequent binding reactions to enhance protein–DNA complex formations.

Competition of specific binding factors from sunflower for the activator element of the 780 gene was achieved by adding various amounts of unlabeled –476/-200 bp fragment from the 780 promoter. A 242 bp *Hpa*II fragment from pUC–19 was added to the binding reaction (already containing the labeled –476/-200 bp fragment) as nonhomologous DNA to compare the specificity of the protein–DNA complexing (Fig. 4–2). Bands "b" and "c" were virtually

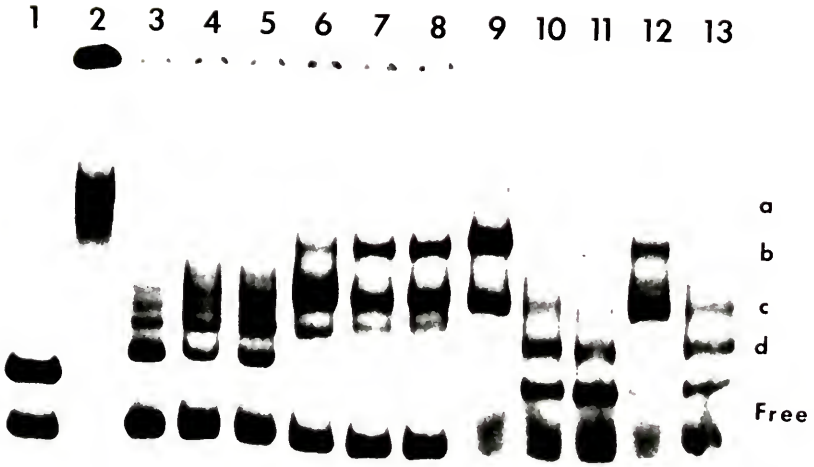


Figure 4–2. Electrophoretic mobility shift assay of the –476/-200 bp 780 gene fragment. The –476/-200 bp fragment was isolated from a 3' deletion clone and 3' end–labeled at the *Eco*R1 site (-476 bp). Approximately 1 ng (5000 cpm) of this fragment was incubated with 2.5 µl of sunflower extract and 5 µg of poly (dI·dC) DNA unless otherwise indicated. The binding reaction salts in all lanes (except lane 3) was 80 mM NaCl, 10 mM Tris–HCl (pH 7.5), 1 mM DTT, 1 mM EDTA and 5% glycerol. Lane **1**, unbound DNA; lane **2**, 2.5 µl extract only; lane **3**, 2.5 µl extract only with no salts; lane **4**, adding 0.2% Triton–X–100; lane **5**, nothing added; Lanes **6** and **9** contain 5 µl extract + 10 µg of poly (dI·dC) DNA without and with the addition of 10 mM $MgCl_2$ + 0.2% Triton–X–100, respectively. Lanes **7** and **8** contain added 10 mM $MgCl_2$ without and with 0.2% Triton–X–100, respectively. Lanes **10–13** contain added 10 mM $MgCl_2$ in each with unlabeled –476/-200 bp fragment of the activator and pUC–19 DNA indicated as follows: **10**, 10 ng of –476/-200 bp fragment; **11**, 100 ng of –476/-200 bp fragment; **12**, 100 ng of 242 bp *Hpa*II fragment from pUC–19 DNA; **13**, 600 ng of 242 bp *Hpa*II fragment from pUC–19 DNA. "a," "b," "c" and "d" refer to the protein–DNA complexes and free refers to the lower unbound band in all lanes. The upper band in lane **1** is a contamination of end-labeled fragment present in the loading dye (10% sucrose, 0.2% bromphenol blue) used for that lane alone.

eliminated by adding 100–fold excess of cold 780 gene –476/-200 bp fragment to the binding reaction (lanes 10 and 11, Fig. 4–2). However, bands "b" and "c" were not affected as much when the same amount of pUC–19 DNA was added (lane 12, Fig. 4–2) demonstrating that this region of the activator element interacted specifically to factors present in the sunflower extract. Band "a" seemed to be a nonspecific complex since it dissappeared when 100 ng of pUC–19 DNA was added.

DNA–binding factors present in another plant species, soybean, were tested for interaction with the activator element (Fig. 4–3). A crude nuclear extract from soybean seedlings was incubated with the –476/-251 bp fragment which contains most of the 780 gene activator element. After separation on a acrylamide gel, a single DNA–protein complex was seen (Fig. 4–3, lane 2). The intensity of this complex was only slightly reduced upon adding as much as 750–fold excess of nonspecific DNA [poly (dI·dC) DNA] in the binding reaction (lane 4). However, this protein–DNA complex was significantly reduced when only 50–fold excess of the cold –476/-251 bp fragment DNA was added with the labeled fragment. DNA from pUC–19 was not able to compete with the activator element for specific binding of these nuclear factors (data not shown). These results indicate that soybean also contains factors which specifically interact with the same portion of the 780 gene activator element that was shown previously to interact with sunflower nuclear factors.

DNase I Protection Analysis of Soybean Factor Interaction with the 780 Activator Element

A DNase I protection analysis was attempted to delineate more accurately regions involved with factor–binding to the activator element. A 3' end–labeled fragment from –476 to –200 bp from the 780 promoter was used in the binding reaction with crude nuclear extracts from soybean. The protein–DNA complex

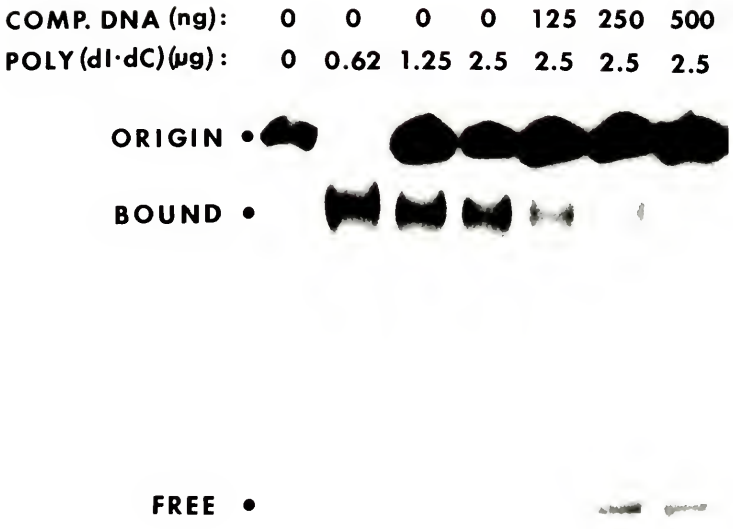


Figure 4–3. Electrophoretic mobility shift assay of a –476/-252 bp fragment from the 780 activator region using soybean extract. Unfractionated nuclear extract (4 µg) from soybean nuclei was incubated with 1 to 2 ng of 5' end–labeled DNA. The DNA used for binding was a 224 bp fragment (-476 to –252 bp) and contained part of the activator element of the 780 promoter. Homologous competitor DNA and poly (dI·dC) DNA was added in varying amounts in the binding reaction as indicated for each lane. Origin refers to the wells of the acrylamide gel, bound is the protein–DNA complex and free is the unbound labeled fragment.

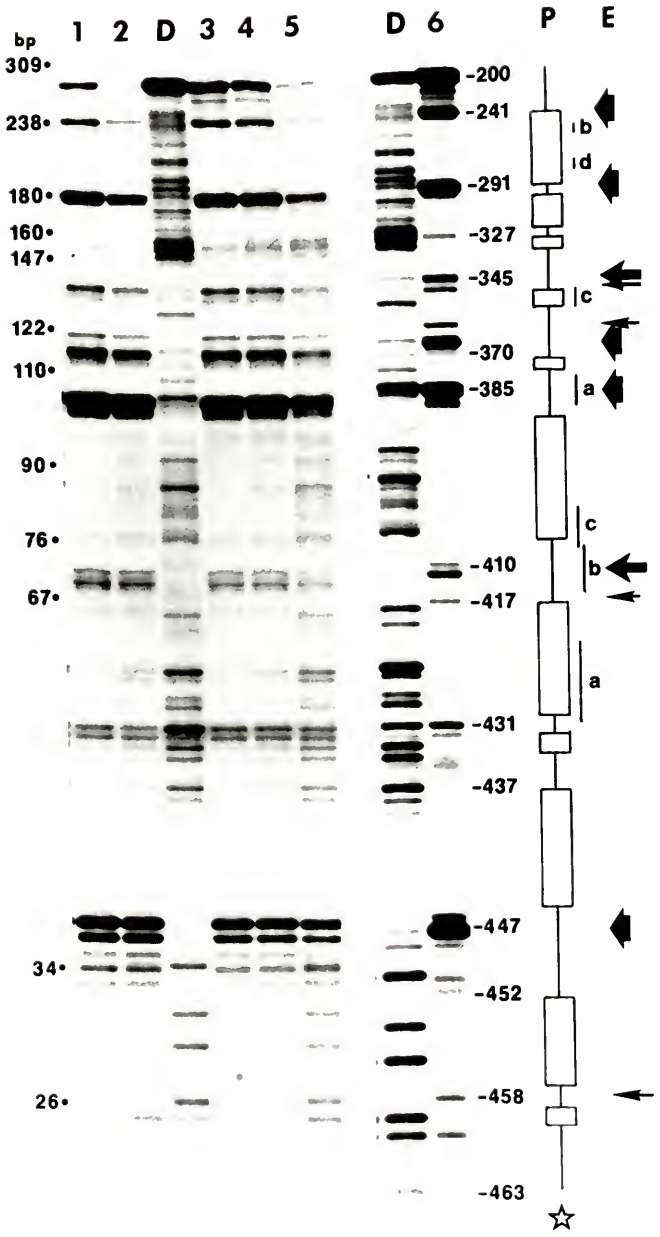
was subsequently digested with limited amounts of DNase I and the labeled DNA analyzed by fractionation on a DNA sequencing gel. Protected regions (very reduced band intensities) of the labeled DNA fragment were seen (lanes 3 and 6, Fig. 4–4) when compared to DNase I generated bands from labeled DNA that was not incubated with extract (lane D, Fig. 4–4). Addition of nonspecific pUC–19 DNA in the binding reaction prior to DNase I digestion did not significantly change the banding pattern (lanes 1 and 2, Fig. 4–4). The protected region, or footprint, was the result of a specific interaction of factors in the extract with the DNA since protection was reduced by the addition of a 200–fold excess of unlabeled 780, –476/–200 bp fragment (lane 5, Fig. 4–4). The protected region encompassed most of the –476/-200 bp fragment which does not coincide with the results of the electrophoretic mobility shift assay involving the –427/-290 bp fragment presented in Fig. 4–1. This discrepancy may be due to the fact that the electrophoretic mobility shift assay involved sunflower extract whereas this DNAse I footprint was generated using soybean extract.

Incubation with the soybean extract also resulted in several DNase I hypersensitive sites (Fig. 4–4, arrows). The intensities of a few of these bands were reduced upon addition of specific competitor DNA suggesting that at least a portion of these sites are the result of specific interaction with nuclear factors.

Conclusion

The activator region of the 780 gene was shown to specifically interact in vitro with crude nuclear extracts from two plant species. This interaction was roughly mapped by the electrophoretic mobility shift assay, and

Figure 4-4. DNase I footprint analysis on the -476/-200 bp 780 activator element. The -476/-200 bp fragment from the 780 promoter was 3' end-labeled at the *Eco*R1 site as in Fig. 4-2. This fragment was incubated with 3 μg of crude soybean extract and 100 ng ΦX174 *Hae*III-digested DNA, and digested with DNase I as described in Materials and Methods unless otherwise indicated. Varying amounts of homologous and nonhomologous DNA was added to the incubation prior to the DNase I digestion also as indicated. Lane D, no extract; lanes 1 and 2, adding 50 and 200 ng of the 242 bp *Hpa*II fragment from pUC-19 DNA, respectively; lanes 3 and 6, nothing added; lanes 4 and 5, adding 50 and 200 ng of unlabeled -476/-200 bp 780 fragment, respectively. P refers to protected regions (open boxes) whereas E refers to the hypersensitive bands (arrows). "a", "b", "c", and "d" indicate the locations of the short repeats found in the 780 promoter. The open star denotes the 3' end-label at -476 bp site. Size markers were determined by a pBR322 *Hpa*II digestion (not shown). Numbers at the right indicate locations of the DNase I-generated bands relative to the start of transcription of the 780 gene.


bp
1
2
D
3
4
5
D
6
P
E
309
238
180
160
147
122
110
90
76
67
34
26
-200
-241
-291
-327
-345
-370
-385
-410
-417
-431
-437
-447
-452
-458
-463
b
d
c
a
c
b
a

examined in more detail by the DNase I protection analysis. The specificity of the protein–DNA complexes was demonstrated by the presence of bands with reduced mobilities that could be reduced by the addition of homologous DNA, but not with nonspecific DNA. Factors from both sunflower and soybean were able to specifically bind to this DNA element suggesting a conservation of factors present in these two species of plants. Although due to a limited number of observations, the 780 gene has only been shown to be transcribed in sunflower and tobacco tissues (73, 168), this gene probably functions in other plants as well since the other T–DNA genes can be expressed in a wide variety of hosts.

The electrophoretic mobility shift assay using the sunflower extract resulted in a complicated pattern of multiple bands (Fig. 4–2). The presence of such multiple bands may indicate multiple factor binding sites in which some of the interactions are specific while others are nonspecific. Others have reported similar patterns of protein–DNA complexes in animal systems (50, 66, 142, 147). Factor binding to the immunoglobulin κ–chain enhancer sequences results in a heterogeneity of binding patterns depending on the tissue–line used for making the nuclear extracts (66, 137). A region of the mouse c–*fos* gene promoter also results in multiple banding patterns when complexed with nuclear factors. In some cases, the additional complexes could not be eliminated with a large excess of homologous or nonhomologous DNA (50); suggesting that certain types of complexes may have discrete gel mobilities, but are the result of some irreversible interactions. There are many possibilities that might cause a multiple banding pattern, such as the existence of different DNA–binding proteins in the same extract, or the binding factor may be present in multimeric forms binding to more than one site.

Nonspecific binding of histones, or similar proteins, all along the length of the DNA fragment may also cause such patterns.

In contrast to the complex banding pattern seen using sunflower extracts, binding of the soybean nuclear extract to a portion of the activator element (-476 to –251 bp) resulted in a single factor–DNA complex. This complex was specific in nature since it could be competed against with homologous DNA but not with nonhomologous DNA. It is possible the sunflower extract may have contained several different factors resulting in multiple bands, whereas the soybean extract contained a single factor producing a single band when bound. However, the presence of a single protein–DNA complex while using the electrophoretic mobility shift assay may still be a result of multermeric binding sites by a particular factor as in the case of the immunoglobulin heavy chain enhancer DNA (3, 137). A fragment containing sequences of this enhancer could form a single band when incubated with nuclear extracts from HeLa and human B lymphoma EW tissue–line; although, assays involving protection from DNase I or methylation demonstrated that the heavy chain enhancer interacted with a factor(s) at multiple sites. Multiple binding also seems to be the case when the DNase I protection analysis was done with the 780 activator element incubated with extract from soybean.

The DNase I protection analysis delineated soybean factor binding sites within the 780 activator region ranging in size from a few nucleotides to approximately 30 bp. Factor binding to most of these regions was specific since only homologous DNA (-476/-200 bp fragment of the 780 promoter) successfully competed for the binding sites. The presence of closely spaced, multiple binding sites were seen in the 780 activator element which can also be found in the DNase I protection analysis of the SV40 promoter and the immunoglobulin heavy chain enhancer (3, 164). In the SV40 promoter,

approximately 5 or 6 closely spaced sites of DNase I protection exist over a region of 250 bp which includes the TATA box, the 21 bp repeats, and the 72 bp enhancer repeat (164). In addition, five sites ranging in sizes from 4 to 48 bp in the immunoglobulin heavy chain enhancer region are evident when protected by factors present in HeLa or BJA–B nuclear extracts from DNase I digestion (3). These results from mammalian studies seems to be similar to the binding pattern of the 780 activator element to factors from soybean extracts.

Some DNase I generated bands were greatly enhanced when factors present in the soybean extract were bound to the activator element. The intensities of most of these DNase I hypersensitive sites were reduced when unlabeled homologous DNA was added to the binding reaction whereas the same sites did not change upon adding nonhomologous DNA. Enhancement of DNase I generated bands due to increased sensitivity to DNase I are also seen in the footprint analysis of the SV40 promoter (164). Some of these bands are also reduced upon adding specific SV40 DNA in the binding reaction while others are unaffected. These hypersensitive bands are thought be a result of the DNA–binding factors interacting with their specific sites and altering conformations of neighboring sequences which makes them more susceptible to DNase I digestion. Another possible cause may involve the bound nuclear factor blocking the DNase I enzyme and causing the enzyme to selectively digest at exposed sites at the boundaries of the protein–DNA complexes. The latter possibility seems to be the case for the soybean factors binding to the activator element since many of these enhanced bands appear between protected areas, leaving only short stretches of DNA exposed for digestion.

Some of the protected regions of the activator element include the short repeats found in the 780 promoter (see Fig. 2–14). The most distal *a* and *c* repeats reside in two of these areas. The second *c* repeat centered at position

–404 bp seems also to be protected from DNase I digestion. Since the 5' deletion to –427 bp did not significantly affect transcription of the 780 gene, the protected regions between –463 to –437 bp may not be as important to promoter activity as regions closer to these repeat motifs. Removal of the 5' flanking region to position –396 bp reduced the level of transcription to 55% (see Fig. 2–11); therefore, the protected regions between –437 and –385 bp undoubtedly contribute to the function of the activator in promoting transcription. Deletion of the remainder of the activator element virtually eliminated transcription to barely detectable levels indicating that the protected regions between –385 to –241 bp also were required for normal function of this element.

It is evident that the activator region of the 780 gene specifically interacts with factors present in crude nuclear extracts from higher plants. Since the activator was previously shown to be required for the 780 gene transcription, and could function upstream independent of position or polarity, the protected sites may be caused by the binding of same factors required for activator function in vivo. This question can not be directly addressed until a reliable plant in vitro transcription system has been developed that will allow reconstitution of purified or partially purified nuclear factors to activate faithful transcription at levels reflecting in vivo systems.

CHAPTER 5

SUMMARY

Promoter domains of the 780 gene were examined by the S1 nuclease mapping using a double gene shuttle vector system developed for this study. The vector system incorporated a copy of the 780 test and reference gene providing an internal standard for precise assessment of each mutation introduced into the test gene promoter. The double gene shuttle vector system has provided a means of quantifying levels of transcription reproducibly and could be applied to other plant genes as well. This vector system has several advantages over other gene expression systems for studying nontissue–specific, constitutively expressed plant genes. One advantage is that transcript levels can be determined directly rather than depend on enzyme function to quantitate expression levels. Secondly, gene copy number discrepancies between the test and reference gene can be avoided since one copy of each is present on the same DNA introduced into the plants. Thirdly, using a homologous reference gene as an internal standard virtually eliminates the possibility of differential expression between heterologous promoters that may occur due to variable growth conditions between groups of pooled tumors. Finally, this *A. tumefaciens* based system results in a low copy number (1–5 per genome); and insertion into the plant genome, facilitates pooling of large populations of separate plant transformation events. The number of tumors pooled was sufficient to minimize variability between

individual tumors since mutant promoter activity between independent experiments varied 10% or less.

Based on the S1 nuclease mapping analysis, two start sites of transcription were found in the 780 gene promoter. These start sites are positioned approximately 30 bp downstream from their respective TATA boxes. The 780 gene, therefore, contains two overlapping promoters, the minor promoter being much less active than the major promoter. Dual transcriptional initiation sites exist in other plant and animal genes such as the mouse thymidilate synthase (27), ovalbumin (133), *Drosophila* Adh (60), soybean nodulin–23 and soybean leghemoglobin genes (99). Alternate start sites are usually differentially utilized as a mechanism for developmental or tissue–specific control. For example, the mouse α_1–amylase gene has two start sites; one is utilized in the salivary glands while the other initiates in liver cells (174). In another example, the two initiation sites of the *Drosophila Adh* gene are differentially regulated during development and may possibly involve a specific nuclear factor (60). Although no evidence exists for differential expression of the 780 gene, the two start sites could be differentially used depending on the tissue or host plant.

Three functional domains in the promoter of the 780 gene were identified and are presented in Fig. 5–1. The schematic presents the 780 5' flanking region as two overlapping promoters corresponding to the major and minor promoter. Both promoters have been subdivided into three major groups, the activator, upstream element and TATA box, based on sequence and function as determined by this study. The most common element, the TATA box, was shown to be essential for transcription of the 780 gene, since removal of the TATA box drastically reduced transcription. When the cap site of the minor promoter was deleted, the start of transcription was still positioned

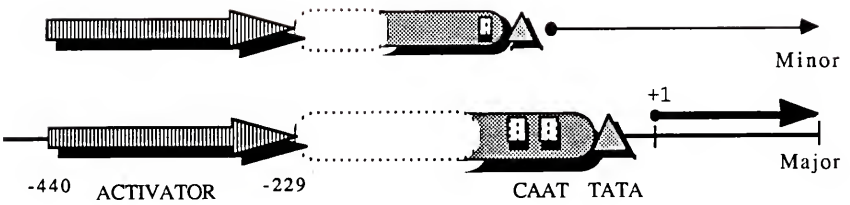


Figure 5–1. Schematic of the major and minor promoter of the 780 gene. The triangles represent the TATA boxes while the small boxes denote positions of the CAAT boxes. The 5' boundary of the upstream element (long–rounded box) is not well defined. The arrow represents the activator element.

approximately 30 bp downstream of the TATA box location demonstrating the role of TATA in positioning transcriptional initiations.

Based on internal deletions, the upstream element was shown to be required under certain conditions, since small disruptions adversely affected transcriptional activity. Yet this element could either be replaced by foreign sequences (Φ X 174), or totally removed, and the 780 gene can still be transcribed efficiently. Although information is very limited, the functional domain of the upstream element in T–DNA genes may be longer than in typical animal genes. This can be seen by the example of the 5' deletions and internal deletions of the AGS gene demonstrating that a region from –168 to –76 bp cannot be disrupted without adversely affecting its transcription (Bandyopadhyay, Bruce, and Gurley, in preparation). Another example is the NOS promoter requiring sequences to –155 bp of the promoter for efficient transcription in tobacco. The CaMV 35S gene also requires sequences upstream of –168 bp for wild type activity. From these few examples it can be argued that the upstream element may actually be ca. –170 to –40 bp in these genes, longer than the average length of the animal gene upstream element which extends from –40 to –110 bp (35). The 780 gene also seems to have an extended upstream element from ca. –200 to –37 bp although the 5' boundary of this element is not well defined.

The third region, the activator element, has enhancer–like qualities as evidenced by its ability to function regardless of polarity and location upstream of a truncated 780 gene. The activator was, however, unable to stimulate transcription while positioned downstream of the test gene. This element was also shown to specifically interact with nuclear factors from two plant species. Since the activator element could function over larger distances than are normally found in the wild type gene, it seems likely that the

mechanism of activator function may involve looping out the intervening DNA. This type of mechanism has been postulated in the case of the λ repressor bound to the operater sites. Approximately 60 to 70 bp can be inserted in between the two adjacent operator sites in the lytic promoter of λ (63). With a repressor dimer binding to each site, the repressors still can interact which each other by bending the DNA in between.

Many more experiments could follow this study in examining the transcriptional control of the 780 gene. Probably the most useful experimental approach in providing accurate information on the promoter substructure would rely on linker–scan mutations. This approach replaces short sequences in the promoter with linker DNA without altering spacing. A large number of linker–scan mutations would accurately delineate the boundaries of internal domains present in the promoter. Following such an analysis, the questions of spacing and helical alignment of distinct promoter domains could then be addressed more carefully. By adding or deleting sequences between promoter elements, more information could be gained concerning the mechinism of transcriptional factors interacting with each other. Such an analysis was previously made on the SV40 promoter by Takahashi et al. (151) in which they introduced linker DNA in various sizes between the TATA, upstream element and the enhancer element. They found that the elements of the SV40 promoter require a stringent alignment to efficiently activate transcription.

In addition, the 780 activator element could be examined more closely by point mutations within the –440 to –229 bp sequence, determining if particular sequences are essential for activator function. Site–directed mutagenesis of the short repeat sequences present in this element or of other regions can conclusively establish functional roles for such sequences. The ability of the activator element to function on heterologous promoters could also be

examined. This analysis would determine if the activator acts as a generalized promoter element, or can only promote activity exclusively with the 780 promoter.

The functional domains of T–DNA promoters probably consist of highly conserved transcriptional control elements. Since the T–DNA promoters have a high degree of conservation in promoter function, these promoters may serve as models for the structure of constitutive promoters in plants. Due to the fact that *Agrobacterium* has such a broad host range, the promoters of T–DNA genes are selected and maintained for functioning in a wide variety of plants. Although these promoters may originate on a bacterial plasmid, such a selection pressure would favor adaption of the most highly conserved and nonspecialized aspects of plant promoter structure and function. The T–DNA genes including the 780 gene could thus provide fundamental information concerning plant transcriptional regulation and establish basic foundations for future constructions of other plant genes.

REFERENCES

1. Adhya, S. and M. Gottesman. 1982. *Cell* **29**:939-944.

2 An, G., P. R. Ebert, B. -Y. Yi, and C. -H. Choi. 1986. *Mol. Gen. Genet.* **203**:245-250.

3. Augereau, P. and P. Chambon. 1986. *EMBO J.* **5**:1791-1797.

4. Banerji, J., S. Rusconi, and W. Schaffner. 1981. *Cell* **27**:299-308.

5. Barker, R., K. Idler, D. Thompson, and J. Kemp. 1983. *Plant Mol. Biol.* **2**:335-350.

6. Benoist, C., and P. Chambon. 1981. *Nature (London)* **276**:633-634.

7. Bergman, Y. 1984. *Proc. Natl. Acad. Sci. USA* **81**: 7041-7045.

8. Beringer, J. E., J. L. Beyon, A. V. Buchanan-Wallaston, and A. W. B. Johnston. 1978. *Nature (London)* **276**:633-634.

9. Bernaerts, M. J., and J. DeLey. 1963. *Nature* **197**:406-407.

10. Bevan, M., R. Barker, A. Goldsbrough, M. Jarvis, T. Kavanagh, and G. Iturriaga. 1986. *Nucl. Acids Res.* **14**:4625-4638.

11. Bevan, M. and M. -D. Chilton. 1982. *J. Mol. Appl. Genet.* **1**:539-546.

12. Bradford, M. M. 1976. *Anal. Biochem.* **72**:248-254.

13. Breatnach, R., and P. Chambon. 1981. *Ann. Rev. Biochem.* **50**:349-383.

14. Buchman, I., F.-J. Marner, G. Schröder, S. Waffenschmidt, and J. Schröder. 1985. *EMBO J.* **4**:853-859.

15. Cassé, F., C. Boucher, J. S. Julliot, M. Michel, and J. Dénarié. 1979. *J. Gen. Microbiololgy* **113**:229-242.

16. Chang, A. C. Y., and S. Cohen. 1978. *J. Bacteriol.* **134**:1141-1156.

17. Chang, C., and E. M. Meyerowitz. 1986. *Proc. Natl. Acad. Sci. USA* **83**:1408-1412.

18. Chilton, M.-D., T. C. Currier, S. K. Farrand, A. J. Bendich, M. P. Gordon, and E. W. Nester. 1974. *Proc. Natl. Acad. Sci. USA* **71**:3672-3676.

19. Cochran, M. D., and C. Weissmann. 1984. *EMBO J.* **3**:2453-2459.

20. Cohen, R., M. Sheffery, and C. Kim. 1986. *Mol. Cell. Biol.* **6**:821-832.

21. Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kédinger, and P. Chambon. *Science* **209**:1406-1414.

22. Currier, T. C., and E. W. Nester. 1976. *J. Bacteriol.* **126**:157-165.

23. Czarnecka, E., L. Edelman, F. Schöffl, and J. L. Key. 1984. *Plant Mol. Biol.* **3**:45-58.

24. Czarnecka, E., W. B. Gurley, R. T. Nagao, L. A. Mosquera, and J. L. Key. 1985. *Proc. Natl. Acad. Sci. USA* **82**:3726-3730.

25. Davidson, B., J. M. Egly, E. R. Mulvihill, and P. Chambon. 1983. *Nature (London)* **301**:680-686.

26. DeCleene, M., and J. DeLey. 1976. *Bot. Rev.* **42**:389-466.

27. Deng, T., L. Dawei, C.-H. Jenh, and L. F. Johnson. 1986. *J. Biol. Chem.* **261**:16000-16005.

28. de Villiers, J., W. Schaffner, C. Tyndall, S. Lupton, and R. Kamen. 1984. *Nature (London)* **312**:242-246.

29. de Villiers, J., and W. Schaffner. 1981. *Nucl. Acids Res.* **9**:6251-6254.

30. Dierks, P., A. van Ooyen, M. Cochran, C. Dobkin, J. Reiser, and C. Weissmann. 1983. *Cell* **32**:695-706.

31. Ditta, G., S. Stanfield, D. Corbin, and D. R. Helinski. 1980. *Proc. Natl. Acad. Sci. USA* **77**:7347-7351.

32. Dudley, R., and A. A. Travers. 1984. *Cell* **38**:391-398.

33. Dynan, W. S., S. Sazer, R. Tjian, and R. Schimke. 1986. *Nature (London)* **319**:246-248.

34. Dynan, W. S., and R. Tjian. 1983. *Cell* **35**:79-87.

35. Dynan, W. S., and R. Tjian. 1985. *Nature (London)* **316**:774-778.

36. Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. Riel, B. G. de Forget, S. W. Weissmann, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot. 1980. *Cell* **21**:653-668.

37. Ellis, J., M. Ryder, and M. Tate. 1984. *Mol. Gen. Genet.* **195**:466-473.

38. Ephrussi, A., G. M. Church, S. Tonegawa, and W. Gilbert. 1985. *Science* **227**:134-140.

39. Epstein, W., and J. R. Beckwith. 1968. *Ann. Rev. Biochem.* **37**:411-436.

40. Farnham, P. J., and R. T. Schimke. 1986. *Mol. Cell. Biol.* **6**:2392-2401.

41. Ferl, R. J. 1985. *Mol. Gen. Genet.* **200**:207-210.

42. Firak, T., and K. N. Subramanian. 1986. *Mol. Cell. Biol.* **6**:3667-3676.

43. Fluhr, R., C. Kuhlemeier, F. Nagy, and N.-H. Chua. 1986. *Science* **232**:1106-1112.

44. Fraley, R. T., S. G. Rogers, R. B. Horsch, P. Sanders, J. Flick, S. Adams, M. L. Bittner, L. Fink, J. S. Fry, G. R. Galluppi, S. B. Goldberg, N. L. Hoffmann, and S. C. Woo. 1983. *Proc. Natl. Acad. Sci. USA* **80**:4803-4807.

45. French, R., M. Janda, and P. Ahlquist. 1986. *Science* **231**:1294-1297.

46. Fromm, M., and P. Berg. 1982. *J. Mol. Appl. Genet.* **1**:457-481.

47. Fujita, T., S. Ohno, H. Yasumitsu, and T. Taniguchi. 1985. *Cell* **41**:489-496.

48. Garcia, J. V., L. Bich-Thuy, J. Stafford, and C. Queen. 1986. *Nature (London)* **322**:383-385.

49. Gidoni, D., W. S. Dynan, and R. Tjian. 1984. *Nature (London)* **312**:409-413.

50. Gilman, M. Z., R. N. Wilson, and R. A. Weinberg. 1986. *Mol. Cell. Biol.* **6**:4305-4316.

51. Gluzman, Y. 1985. *Eukaryotic Transcription: the role of cis- and trans-acting elements in initiation* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.).

52. Graves, B., P. F. Johnson, and S. L. McKnight. 1986. *Cell* **44**:565-576.

53. Green, M. R., R. Treisman, and T. Maniatis. 1983. *Cell* **35**:137-148.

54. Grosschedl, R., and M. L. Birnstiel. 1980. *Proc. Natl. Acad. Sci. USA* **77**:1432-1436.

55. Gurley, W. B., E. Czarnecka, R. Nagao, and J. L. Key. 1986. *Mol. Cell. Biol.* **6**:559-576.

56. Gurley, W. B., J. D. Kemp, M. J. Albert, D. Sutton, and J. Callis. 1979. *Proc. Natl. Acad. Sci. USA* **76**:2828-2832.

57. Hagan, G., and T. Guilfoyle. 1985. *Mol. Cell. Biol.* **5**:1197-1203.

58. Hayashi, S., and H. Kondoh. 1986. *Mol. Cell. Biol.* **6**:4130-4132.

59. Hayashi, S., H. Kondoh, K. Yasuda, G. Soma, Y. Ikawa, and T. S. Okada. 1985. *EMBO J.* **4**:2201-2207.

60. Heberlein, U., B. England, and R. Tjian. 1985. *Cell* **41**:965-977.

61. Herr, W., and Y. Gluzman. 1985. *Nature (London)* **313**:711-714.

62. Hochschild, A. H., N. Irwin, and M. Ptashne. 1983. *Cell* **32**:319-325.

63. Hochschild, A. H., and M. Ptashne. 1986. *Cell* **44**:681-687.

64. Hooykaas-van Slogteren, G. M. S., P. J. J. Hooykaas, and R. A. Schilperoort. 1984. *Nature (London)* **311**:763-764.

65. Horsch, R., J. Fry, N. Hoffmann, N. Eichholtz, S. Rogers, and R. Fraley. 1985. *Science* **227**:1229-1231.

66. Hromas, R., and B. V. Ness. 1986. *Nucl. Acids Res.* **14**:4837-4848.

67. Hutchison, C. A., S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, and M. Smith. 1978. *J. Biol. Chem.* **253**:6551-6560.

68. Jackson, M., and J. Ingle. 1973. *Plant Physiol.* **51**:412-414.

69. Jensen, J. S., K. A. Marcker, L. Otten, and J. Schell. 1986. *Nature (London)* **321**:669-674.

71. Jones, K. A., and R. Tjian. 1985. *Nature (London)* **317**:179-182.

70. Jones, K. A., K. Yamamoto, and R. Tjian. 1985. *Cell* **42**:559-572.

72. Kadesch, T., and P. Berg. 1986. *Mol. Cell. Biol.* **6**:2593-2601.

73. Karcher, S. J., V. J. DiRita, and S. B. Gelvin. 1984. *Mol. Gen. Genet.* **194**:159-165.

74. Karin, M., A. Hoslinger, H. Holtgreve, R. I. Richards, P. Krauter, H. M. Westphal, and M. Beato. 1984. *Nature (London)* **308**:513-519.

75. Kaulen, H., J. Schell, and F. Kreuzaler. 1986. *EMBO J.* **5**:1-8.

76. Kay, R. J., R. J. Boissy, R. H. Russnak, and E. P. M. Candido. 1986. *Mol. Cell. Biol.* **6**:3667-3676.

77. Keil, M., J. Sanchez-Serrano, J. Schell, and L. Willmitzer. 1986. *Nucl. Acids Res.* **14**:5639-5640.

78. Klee, H., A. Montoya, F. Horodyski, C. Lichtenstein, D. Garfinkel, S. Fuller, C. Flores, J. Peschon, E. Nester, and M. Gordon. 1984. *Proc. Natl. Acad. Sci. USA* **81**:1728-1732.

79. Klein, S., F. Sablitzky, and A. Radbruch. 1984. *EMBO J.* **3**:2473-2476.

80. Kohno, K., M. Sullivan, and Y. Yamada. 1985. *J. Biol. Chem.* **260**:4441-4447.

81. Koncz, C., and J. Schell. 1986. *Mol. Gen. Genet.* **204**:383-396.

82. Koncz, C., H. DeGreve, D. André, F. Deboeck, M. Van Montague, and J. Schell. 1983. *EMBO J.* **2**:1597-1603.

83. Kovacs, B., and P. Butterworth. 1986. *Nucl. Acids Res.* **14**:2429-2442.

84. Kriegler, M., and M. Botchan. 1983. *Mol. Cell. Biol.* **3**:325-339.

85. Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss. 1982. *Proc. Natl. Acad. Sci. USA* **79**:6453-6457.

86. Lamppa, G. K., G. Morelli, and N.-H. Chua. 1985. *Mol. Cell. Biol.* **5**:1370-1378.

87. Larsen, K., and B. U. Jochimsen. 1986. *EMBO J.* **5**:15-19.

88. Lawrence, C. W., F. Sherman, M. Jackson, and R. Gilmore. 1975. *Genetics* **81**:615-629.

89. Lentwiler, L., E. M. Meyerowitz, and E. M. Tobin. 1986. *Nucl. Acids Res.* **14**:4051-4064.

90. Lewis, M. K., and R. R. Burgess. 1982. Eucaryotic RNA polymerases. In *The Enzymes*, 3rd ed., vol **15**, B. P. Boyer, ed. Academic Press, New York, N. Y., pp. 109-153.

91. Lichtenstein, C., H. Klee, A. Montoya, D. Garfinkel, S. Fuller, C. Flores, E. Nester, and M. Gordon. 1984. *J. Mol. Appl. Genet.* **2**:354-362.

92. Lis, J., J. Simon, and C. Sutton. 1983. *Cell* **35**:403-410.

93. Luciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Capecchi. 1983. *Cell* **33**:705-716.

94. Lusky, M., L. Berg, H. Weiher, and M. Botchan. 1983. *Mol. Cell. Biol.* **3**:1108-1122.

95. Maniatis, T., E. Fritsch, and J. Sambrook. 1982. *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.).

96. Marchuk, D., S. McCrohon, and E. Fuchs. 1985. *Proc. Natl. Acad. Sci. USA* **82**:1609-1613.

97. Mathis, D., and P. Chambon. 1981. *Nature (London)* **290**:310-315.

98. Matzke, A. J. M., and M.-D. Chilton. 1981. *J. Mol. Appl. Genet.* **1**:39-49.

99. Mauro, V. P., T. Nguyen, P. Katinakis, and D. P. S. Verma. 1985. *Nucl. Acids Res.* **13**:239-249.

100. Maxam, A., and W. Gilbert. 1980. *Methods Enzymol.* **65**:499-560.

101. McGrogan, M. M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke. 1985 *J. Biol. Chem.* **260**:2307-2314.

102. McKnight, S. L. 1982. *Cell* **31**:355-365.

103. McKnight, S. L., and R. Kingsbury. 1982. *Science* **217**:316-324.

104. Melton, D. W., C. McEwan, A. B. McKie, and A. M. Reid. 1986. *Cell* 44:319-328.

105. Mercola, M., J. Goverman, C. Mirell, and K. Calame. 1985. *Science* **227**:266-270.

106. Merlo, D., R. Nutter, A. Montoya, D. Garfinkel, M. Drummond, M.-D. Chilton, M. Gordon, and E. Nester. 1980. *Mol. Gen. Genet.* **177**:637-643.

107. Miller, J. H. 1972. *Experiments In Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.).

108. Moreau, P., R. Hen, B. Wasylyk, R. Everrett, M. P. Gaub, and P. Chambon. 1981. *Nucl. Acids Res.* **9**:6047-6069.

109. Morelli, G., F. Nagy, R. T. Fraley, S. G. Rogers, and N.-H. Chua. 1985. *Nature (London)* **315**:200-204.

110. Muria, N., and J. D. Kemp. 1982. *Proc. Natl. Acad. Sci. USA* **79**:86-90.

111. Muria, N., D. W. Sutton, M. G. Murray, J. L. Slightom, D. J. Merlo, N. A. Reichert, C. Sengupta-Gopalan, C. A. Stock, R. F. Barker, J. D. Kemp, and T. C. Hall. 1983. *Science* **222**:476-482.

112. Myers, R., K. Tilly, and T. Maniatis. 1986. *Science* **232**:613-618.

113. Nagao, R. T., E. Czarnecka, W. B. Gurley, F. Schöffl, and J. L. Key. 1985. *Mol. Cell. Biol.* **5**:3417-3428.

114. Nester, E. W., M. P. Gordon, R. M. Amasino, and M. F. Yanofsky. 1984. *Annu. Rev. Plant Physiol.* **35**:387-413.

115. Neuhaus, G., P. Neuhaus-Url, P. Gruss, and H. G. Schweiger. 1984. *EMBO J.* **3**:2169-2172.

116. Norrander, J., T. Kempe, and J. Messing. 1983. *Gene* **26**:101-106.

117. Odell. J., F. Nagy, and N.-H. Chua. 1985. *Nature (London)* **313**:810-812.

118. Ooms, G., P. M. Klapwijk, J. A. Poulis, and R. A. Schilperoort. 1980. *J. Bacteriol.* **144**:82-91.

119. Ott, M.-O., L. Sperling, P. Herbomel, M. Yaniv, and M. C. Weiss. 1984. *EMBO J.* **3**:2505-2510.

120. Otten, L., H. DeGreve, J. Hernalsteens, M. Van Montague, O. Schrieder, J. Straub, and J. Schell. 1981. *Mol. Gen. Genet.* **183**:209-213.

121. Parker, C. S., and J. Topol. 1984. *Cell* **36**:357-369.

122. Paul, A.-L., V. Vasil, I. Vasil, and R. J. Ferl. 1986. *Proc. Natl. Acad. Sci. USA* (in press).

123. Petit, A., S. Delhaye, J. Tempé, and G. Morel. 1970. *Physiol. Veg.* **8**:205-213.

124. Reynolds, G. A., S. K. Basu, T. S. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey. 1984. *Cell* **38**:275-285.

125. Richards, R. I., A. Heguy, and M. Karin. 1984. *Cell* **37**:263-272.

126. Rochester, D., J. Winter, and D. M. Shah. 1986. *EMBO J.* **5**:451-458.

127. Rothstein, S. J., R. A. Jorgensen, K. Postle, and W. S. Reznikoff. 1980. *Cell* **19**:795-805.

128. Salomon, F., R. Deblaere, J. Leemans, J. Hernalsteens, M. Van Montague, and J. Schell. 1984. *EMBO J.* **3**:141-146.

129. Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchinson, P. M. Slocombe, and M. Smith. 1977. *Nature (London)* **265**:687-695.

130. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1981. *J. Mol. Biol.* **143**:161-178.

131. Sassone-Corsi, P., J. Corden, C. Kédinger, and P. Chambon. 1981. *Nucl. Acids Res.* **9**:3941-3958.

132. Sazer, S., and R. Schimke. 1986. *J. Biol. Chem.* **261**:4685-4690.

133. Schibler, U., O. Hagenbüchle, P. K. Wellauer, and A. C. Pittet. 1983. *Proc. Natl. Acad. Sci. USA* **78**:879-883.

134. Schöler, H. R., and P. Gruss. 1984. *Cell* **36**:403-411.

135. Searle, P. F., G. W. Stuart, and R. D. Palmiter. 1985. *Mol. Cell. Biol.* **5**:1480-1489.

136. Sederoff, R., A.-M. Stomp, W. S. Chilton, and L. W. Moore. 1986. *Bio/Technology* **4**:647-649.

137. Sen, R. and D. Baltimore. 1986. *Cell* **46**:705-716.

138. Serfling, E., M. Jasin, and W. Schaffner. 1985. *Trends Genet.* **1**:224-230.

139. Shaul, Y., W. J. Rutter, and O. Lamb. 1985. *EMBO J.* **4**:427-430.

140. Shaw, C. H., G. H. Carter, M. D. Watson, C. H. Shaw. 1984. *Nucl. Acids Res.* **12**:7831-7846.

141. Shuey, D. J., and C. S. Parker. 1986. *J. Biol. Chem.* **261**:7934-7940.

142. Shuey, D. J., and C. S. Parker. 1986. *Nature (London)* **323**:459-461.

143. Siebenlist, U., R. B. Simpson, and W. Gilbert. 1980. *Cell* **20**:269-281.

144. Silflow, C. D., J. R. Hammett, and J. L. Key. 1979. *Biochem.* **18**:2725-2731.

145. Simpson, J., J. Schell, M. Van Montague, and L. Herrera-Estrella. 1986. *Nature (London)* **323**:551-554.

146. Simpson, J., M. Timko, A. Cashmore, J. Schell, M. Van Montague, and L. Herrera-Estrella. 1985. *EMBO J.* **4**:2723-2729.

147. Singh, H., R. Sen, D. Baltimore, and P. A Sharp. 1986. *Nature (London)* **319**:154-158.

148. Southern, E. M. 1975. *J. Mol. Biol.* **98**:503-517.

149. Southgate, R., A. Ayme, and R. Voellmy. 1983. *J. Mol. Biol.* **165**:35-57.

150. Spena, A., R. Hain, U. Ziervogel, H. Saedler, and J. Schell. 1985. *EMBO J.* **4**:2739-2743.

151. Takahashi, K., M. Vigneron, H. Matthes, A. Wildeman, M. Zenke, and P. Chambon. 1986. *Nature (London)* **319**:121-126.

152. Thomashow, M., R. Nutter, A. Montoya, M. Gordon, and E. Nester. 1980. *Cell* **19**:729-739.

153. Timko, M., A. Kausch, C. Castresana, J. Fassler, L. Herrera-Estrella, G. Van den Broeck, M. Van Montague, J. Schell, and A. Cashmore. 1985. *Nature (London)* **318**:579-582.

154. Tokunaga, K., S. Hirose, and Y. Suzuki. 1984. *Nucl. Acids Res.* **12**:1543-1558.

155. Topol, J., D. Ruden, and C. S. Parker. 1985. *Cell* **42**:527-537.

156. Treisman, R., M. R. Green, and T. Maniatis. 1983. *Proc. Natl. Acad. Sci. USA* **80**:7428-7432.

157. Van Onckelen, H., E. Prinsen, D. Inzé, P. Rüdelshein, M. Van Lijsebettens, A. Follins, J. Schell, M. Van Montague, and J. DeGreef. 1986. *FEBS Let.* **198**:357-360.

158. Walker, M. D., T. Edlund, A. M. Boulet, and W. J. Rutter. 1983. *Nature (London)* **306**:557-561.

159. Wasylyk, B., R. Derbyshire, A. Guy, D. Molko, A. Roget, R. Téoule, and P. Chambon. 1980. *Proc. Natl. Acad. Sci. USA* 77:7024-7028.

160. Wasylyk, B., C. Kédinger, J. Corden, O. Brison, and P. Chambon. 1980. *Nature (London)* **285**:367-373.

161. Wasylyk, B., C. Wasylyk, P. Augereau, and P. Chambon. 1983. *Cell* **32**:503-514.

162. Wei, R., H. Wilkinson, K. Pfeifer, C. Schnieder, R. Young, and L. Guarente. 1986. *Nucl. Acids Res.* **14**:8183-8188.

163. Weiher, H., M. König, and P. Gruss. 1983. *Science* **219**:626-631.

164. Wildeman, A. G., M. Zenke, C. Shatz, M. Winterith, T. Grunström, H. Matthes, K. Takahashi, and P. Chambon. 1986. *Mol. Cell. Biol.* **6**:2098-2105.

165. Willmitzer, L., P. Dhaese, P. H. Schreier, W. Schmalenbach, M. Van Montague, and J. Schell. 1983. *Cell* **19**:729-739.

166. Willmitzer, L., L. Otten, G. Simons, W. Schmalenbach, J. Schröder, G. Schröder, M. Van Montague, and J. Schell. 1981. *Mol. Gen. Genet.* **182**:255-262.

167. Willmitzer, L., G. Simons, and J. Schell. 1982. *EMBO J.* **1**:139-146.

168. Winter, J. A., R. L. Wright, and W. B. Gurley. 1984. *Nucl. Acids Res.* **12**:2391-2406.

169. Wu, C. 1984. *Nature (London)* **311**:81-84.

170. Wu, C. 1984. *Nature (London)* **309**:229-234.

171. Wu, C. 1985. *Nature (London)* **317**:84-87.

172. Wu, H.-M., and D. M. Crothers. 1984. *Nature (London)* **308**:509-513.

173. Yamada, T., C. J. Palm, B. Brooks, and T. Kosuge. 1985. *Proc. Natl. Acad. Sci. USA* **82**:6522-6526.

174. Young, R. A., O. Hagenbüchle, and U. Schibler. 1981. *Cell* **23**:451-458.

175. Zambryski, P., A. Depicker, K. Kruger, and H. M. Goodman. 1982. *J. Mol. Appl. Genet.* **1**:361-370.

176. Zenke, M., T. Grundström, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon. 1986. *EMBO J.* **5**:387-397.

BIOGRAPHICAL SKETCH

Wesley B. Bruce was born November 11, 1959, in Lafayette, Louisiana. He attended Terry Parker High School in Jacksonville, Fl. then was admitted to the University of Florida as a freshman in August of 1978. After receiving a Bachelor of Science degree in microbiology in August of 1981 he immediately began graduate school at the Department of Microbiology and Cell Science under the supervision of Dr. William B. Gurley. Upon receiving his Ph.D., he will join Dr. Peter Quail's laboratory at the University of Wisconsin at Madison, Wisconsin, as a research associate and will continue studies concerning molecular genetics of crop plants.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William B. Gurley, Chairman
Associate Professor of Microbiology
and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Francis C. Davis, Jr., Cochairman
Associate Professor of Microbiology
and Cell Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James Preston
Professor of Microbiology and Cell
Science

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert J. Ferl
Associate Professor of Botany

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

L. Curtis Hannah

L. Curt Hannah
Professor of Horticulture Science

This dissertation was submitted to the Graduate Faculty of the College of Agriculture and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

May 1987

Jack L. Fry

Dean, College of Agriculture

Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08666 983 4